Matters of Time Directionality in Classical Physics[*]

Jean-Christophe Lindner
Department of Physics, Université de Montréal
Montréal, QC, Canada

Updated 24 February 2020

Abstract

This report offers a modern perspective on the question of time directionality as it arises in a semi-classical context, based on key developments in the field of gravitational physics. Important clarifications are achieved regarding, in particular, the concept of time reversal and that of negative energy state. The conditions imposed by the Leibnizian constraint of relational definition of physical attributes is thoroughly examined and significant consequences of applying this consistency requirement are derived. From this analysis emerges an improved understanding of the general relativistic concept of stress-energy of matter as being a manifestation of local variations in the energy density of zero-point vacuum fluctuations. Based on those developments a set of axioms is proposed that enables the derivation of generalized gravitational field equations which actually constitute a simplification of relativity theory in the presence of negative energy matter and vacuum energy. Those results are then applied to provide significant new insights into many aspects of the semi-classical theory of black hole thermodynamics and to offer original solutions to several long-standing problems in theoretical cosmology, including the problem of the nature of dark matter and dark energy, that of the origin of thermodynamic time asymmetry and several other issues traditionally approached using inflation theory.

---

[*]Preprint addresses: arxiv.org/abs/1608.01531, philpapers.org/rec/LINMOT-4.

Contents

List of Figures

List of Tables

Introduction

The reflection which gave rise to the developments that will be introduced in this report started with a very simple question: could gravitation be a repulsive force under certain circumstances and what would it mean for gravitational mass to be negative? Even though there appears to be important difficulties associated with the possibility that a gravitationally repulsive body may exist, particularly in the context of a general relativistic theory, the idea of a symmetry which would have to do with the sign of mass or energy is certainly quite appealing aesthetically. Indeed, if the electric charge and all the other charges turning up in particle physics are allowed to be both positive and negative, why should mass or energy be restricted to positive values? What I came to realize through a careful analysis of the assumptions behind the common idea that gravitationally repulsive matter does not exist is that there is actually a general misunderstanding surrounding the whole idea of negative energy in modern physical theory and that this is the single most important stumbling block that is preventing necessary progress to be achieved in several fields of fundamental theoretical physics. The objective of this essay is to clear up the misunderstanding and to provide a detailed account of the most crucial advances which are made possible by adopting a more consistent approach regarding some essential concepts related to time directionality and their relationships with our current classical theory of gravitation.

I will therefore begin by revisiting the old problem of negative energy states and by explaining the difficulties which arise in the context of the current conception of negative mass. This will allow me to achieve a more consistent integration of the concept of negative energy matter to the classical theory of gravitation by drawing on the analogy provided by the gravitational dynamics of voids in a matter distribution. I will show that traditional expectations regarding the interaction of negative energy matter with itself and

with positive energy matter are inappropriate, because they violate the requirement that all physical properties be defined in a relational way. From this analysis will emerge an improved understanding of the notion of gravitational repulsion involving negative energy matter as a form of dark matter whose existence must be considered unavoidable from both a theoretical and an empirical viewpoint. An alternative set of axioms which allows an appropriate and at last consistent integration of negative energy states to physical theory will then be proposed. I will conclude this portion of my analysis with a reformulation of the relativistic gravitational field equations that provides the foundation for the first-ever bi-metric theory of gravitation that is truly symmetric under exchange of positive and negative energy states and which actually simplifies the original theory in the presence of a non-zero cosmological constant.

What allowed me to achieve a better understanding of the concept of negative energy matter is the acknowledgement that there must exist a fundamental time-direction degree of freedom independent from the thermodynamic concept of time direction. In such a context it emerges that only the sign of energy defined in relation to a given direction of propagation in time is significant from a gravitational viewpoint. Once the significance of this insight was properly assimilated it became possible to develop an alternative concept of time reversal that allows a reformulation of the discrete symmetry operations and a more consistent description of the changes occurring under a reversal of space- and time-related parameters. In order to achieve full consistency, it was necessary to introduce an additional set of discrete symmetry operations of a kind which had never been considered and which transforms a positive energy state into various negative energy states. Those developments then allowed the derivation of an exact binary measure for the entropy of the matter contained within the event horizon of a black hole that reproduces the results of the semi-classical theory in the case of elementary (Planck mass) black holes.

As a consequence of the relatively long period of gestation during which the mere intuitive insights from which this work originates evolved into a revised classical theory of gravitation, I was able to explore the consequences of some of the most decisive results which were reached in the course of that process on a rather large number of questions of fundamental interest. Thus, I can now provide a complete account of the implications of this improved understanding of gravitational physics for classical cosmology theory and in the process achieve a better understanding of several issues related to time

directionality. I will, in particular, provide significant new insight regarding the whole question of dark energy and dark matter and the related problem of the formation of large scale structures. Still by making use of the results derived in the first portion of this report, I will then propose alternative solutions to some outstanding problems of classical cosmology which were originally addressed using inflation theory. I will conclude this analysis by providing a definitive solution to the problem of the origin of time irreversibility which relies on a more accurate estimation of the measures of entropy associated with the gravitational field and the microscopic structure of physical space.

Motivations

It must be mentioned that even though I became interested in the idea underlying the developments discussed in this report based on mostly aesthetic motives, the actual reasons that later fueled my interest in developing a viable model around it were of a more pragmatic nature. In particular, I saw the need that existed, but that few authors recognized, to reformulate the current classical theory of gravitation in a way that would be consistent with the possibility for elementary particles to be found in the negative energy states allowed by special relativistic quantum theories. Indeed, I had come to understand that the current interpretation of negative energy states as merely being those of antiparticles whose behavior is identical to that of ordinary matter from a gravitational viewpoint, was dependent on the a priori assumption that only some of those energy states were allowed. In other words, we had solved the puzzling problem of the prediction of negative energy states by postulating that those states were not allowed, without justifying this very assumption. But if we recognize that the whole spectrum of energy states predicted to exist by quantum theory can in effect be occupied, even if transitions between positive and negative energy states may not be allowed, then we need a classical theory of gravitation that is consistent with this requirement. However, further considerations indicated that the general theory of relativity is not entirely compatible with an appropriate notion of negative energy obeying certain theoretical requirements which must be imposed in order to achieve consistency.

Despite those difficulties I believe that the imperative to provide an appropriate description of negative energy matter should prevail over our will-

ingness to leave untouched the current theory of gravitation, because I have recognized the inadequacy of the arguments against the physical nature of negative energy states, while I also understand that quantum theory constitutes a more appropriate basis to decide what states are allowed for elementary particles. Thus, I persisted in seeking to achieve this integration and as it turned out this insistence was vindicated given that I was able to develop an alternative framework that merely generalizes relativity theory in a very elegant manner without affecting its basic mathematical structure, while allowing an appropriate description of negative energy matter.

But I was also motivated by the desire to obtain a better agreement between theoretical predictions and astronomical observations concerning certain aspects of the gravitational dynamics of the universe. In particular, there was the exceptionally severe disagreement between most theoretical derivations of the expected value of vacuum energy density and observational constraints on the upper (positive or negative) value of the cosmological constant. Very early on I saw that the hypothesis that matter should be allowed to exist in a negative energy state could potentially provide a whole new class of contributions to zero-point vacuum energy which would be the exact opposite of those already considered in conventional calculations and which could naturally allow an overall cancellation of all contributions if some level of symmetry exists between the viewpoint of positive energy observers and that of negative energy observers. Here again I chose not to ignore, as most people did, what seemed to be the necessary conclusion that matter must be allowed to occupy the currently forbidden negative energy states if we are to obtain a compensation for the known contributions to vacuum energy. Despite the apparent difficulties, perceived or real, associated with negative energy as a possible state of matter it had become very clear to me that this was a hypothesis which had become unavoidable.

Finally, I also wanted to bring some much needed clarity to the theoretical context in which we are to address the problem of the elaboration of a theory of the gravitational interaction compatible with the basic principles of quantum theory. Here I will show the essential role played by the discrete spacetime and momentum-energy symmetry operations (appropriately redefined and extended to comply with an improved concept of time reversal) in characterizing states of matter at the spatial scale and energy level at which we can expect the gravitational interaction among elementary particles to be as strong as the other known interactions. This will be achieved by demonstrating the relevance of those symmetry operations for a definition of the

microstates that must be taken into consideration in order to provide an appropriate measure of black hole entropy.

Approach

Basically the approach I will follow consists in explaining how some specific aspect of the quantum world, namely the ignored possibility for both positive and negative energy states to propagate forward and backward in time, changes our understanding of the classical theory of gravitation and allows to actually improve and simplify its formulation in a way that will have decisive consequences for the description of certain phenomena which are taking place on the cosmological scale. The level of this discussion is clearly philosophical, but remains very precise in its reference to quantitative aspects and concepts, unlike most philosophical essays concerning physics. Mathematical developments will be kept to a bare minimum, however, and will be introduced only when absolutely necessary and of utmost significance. This is obviously in contrast with the current tendency observed in the physical sciences to focus on technical aspects and to relegate epistemology to the backseat.

Concerning the methodology which is reflected in the style of this report I must emphasize that I have been introduced to quantitative methods very early on, but I later came to realize that in the context where all the really useful mathematical developments that could be carried out in the field of fundamental theoretical physics have already been performed over and over again by competent people, real progress can only arise at the level of interpretation. Indeed, a fully consistent interpretation of the existing frameworks is currently missing, perhaps because the vast majority of competent researchers prefer to dedicate their efforts to more technical aspects, and this is restraining our ability to distinguish between what are viable developments and what is logically and empirically inappropriate. But as I do believe that the objective of a philosophy of science should be to explain and to justify, through logical arguments constrained by observational data, a particular vision of the world, and as I'm convinced that it is only when this goal is successfully achieved that we are allowed to consider some vision of the world as a valid representation of it, then this is the objective toward which I directed my efforts.

Furthermore, it is important to note that if mathematical developments do not dominate the content of this report it is also simply a consequence of

the fact that while I have achieved a crucial revision and a necessary improvement of the mathematical framework and of the interpretation of relativity theory in a semi-classical context, I nevertheless ended up confirming the general validity of the basic mathematical structure of the current theory, within a certain limit, so that practically no further mathematical developments were required. The reader must be warned, however, that the density of significant information that is to be found in the text of this document is very high. In some cases, it took me years of dedicated reflection and careful investigation to gain confidence in the validity and inevitability of certain specific results which may be mentioned only once in the main portion of the report, as otherwise the length of the treatise would be excessive. Therefore, you must pay attention to every detail of the discussion and be careful not to miss some important information that may be necessary later on for understanding and appreciating the value of other elements of the discussion. I know that this may sound obvious, but here the difficulty may be so great that it is nevertheless appropriate to issue such a warning. This, however, does not mean that the present report is actually difficult to read, to the contrary. In fact, I tend to follow a rather educational approach according to which I do not avoid making statements and providing explanations, even when they may appear obvious to some or even most readers, because I think that it is better to make too many unnecessary statements than to more or less willingly avoid making some which would have been useful. This approach should not be considered as condescending or as an indication that this work is intended mainly for a beginner audience.

I must mention that I do recognize that the approach I followed in order to achieve the valuable results that will be described and justified here *is* different from that which is usually followed in theoretical physics. But, even if I would not myself have believed that one could achieve such significant results following that kind of method when I started studying physics, which I did the usual way by learning about the mathematics of quantum theory, statistical mechanics and relativity theory, it is through experience and by force of circumstance (although not as a result of mere incompetence), after having slowly and partly unwillingly deviated from the traditional path, that I began to understand that there is real value in such an approach which I developed by making systematic a learning process that initially appeared to merely be a faithful but irresponsible time-wasting improvisation. If the reader is willing to immerse herself in the same experience and loosen her grip on more traditional ways of achieving deep understanding, while nevertheless

being ready to spend considerable efforts to follow rigorous logical reasoning and analysis, I can assure her that she will not be deceived and will learn useful physics, which is not so bad already by today's standards.

Historical context

There are many similarities between the current state in which science finds itself and those through which it went at other crucial turning points in its history. Indeed, the situation we have now arrived at is characterized by an accumulation of unanswered questions which creates an impasse that prevents further progress from being achieved. It is my belief that answering just a few key questions among those will release a great deal of pressure that will greatly facilitate future theoretical research. When we examine the present situation in physics it becomes clear in effect that if there are questions which we are justified in not being able to answer right now, because they are related to what may be said of reality under conditions which we cannot yet reproduce in experiments (think of trying to explain the origin of the free parameters of the standard model of particle physics), there are also questions which have to do with known difficulties which we have puzzled about for a long time and which we have no reason to believe further experiments may be particularly useful in helping resolve. But those are problems whose existence is often simply unknown to most people or which are sometimes considered to have already been solved, while careful examination shows that this is not always entirely the case. Most current programs in fundamental theoretical physics are focused on trying to solve the problems raised by questions of the first type and this is unfortunate, because here is precisely the domain in which progress is limited by technological constraints of a practical nature and the cost of achieving the required experiments. Very early on I recognized that if I was going to enable progress to be made in physics I had to concentrate on questions of the latter type, where progress could occur not only in my lifetime, but also as a consequence of the success or failure of my own enterprise.

Among the questions we may have hope to answer using our current knowledge is the question I mentioned earlier on as having being that which launched the reflection process from which this report emerged. It is in effect one of those unsolved questions whose very existence is usually unrecognized or which is considered to have already been solved, while this is clearly not the

case, as I will explain later. You will not see it mentioned in most accounts as being one of today's open questions in physics, but it is one of the most important categories of question regarding classical physics and a field most people currently consider to be free of major difficulties. This problem of negative energy states could actually be called the 'classical gravitation theory problem' or the 'general relativity problem', because properly answering that question requires introducing slight modifications to that theory, which actually consists in a generalization of its own founding principles. This is the first question I will address in this report and satisfactory answers will be provided to the mostly unrecognized issues it currently raises. Doing so will require reconsidering the significance of certain aspects of the problem of vacuum energy and gaining a new understanding of the gravitational effects of homogeneous and inhomogeneous matter distributions that can be extended to our description of the physical vacuum.

An additional category of questions which is also related to classical gravitation theory can be collectively described as the 'cosmology problem'. It asks what is the origin of the constants of the standard model of cosmology, what is dark matter and what is dark energy, how are we to resolve the flatness and horizon problems, and what explains the existence of the largest structures in which clusters of galaxies are observed to be organized? It also asks why it is that the energy which is contained in zero-point vacuum fluctuations is so low in comparison with the very large value that is provided by most theoretical estimates? Finally it asks whether there was a beginning to time in the past and how matter was created during the Big Bang? While it is often considered that some of those questions have already been answered by developments like inflation theory, I will explain that there remain important unresolved issues in this context and that we are justified in seeking alternative answers, which I will show do exist. In fact, even though the objectives I had in mind when I started this research project were quite humble, in the end I was able to provide original solutions to nearly all aspects of the cosmology problem.

But I will also address a further category of questions that is usually considered to regard classical physics, but which actually sits right at the interface between the classical theory of gravitation and quantum theory. This the traditional question of the origin of the statistical properties of matter which are reflected in the unidirectionality of the evolution in time of systems with a large number of degrees of freedom. Given that this problem of the origin of the thermodynamic arrow of time can be traced back to the

peculiar properties of the distribution of matter energy which existed during the first instants of the Big Bang (as I will explain), it follows that the question of the origin of the unidirectionality of thermodynamic processes is in effect also a question for cosmology and as such it will hugely benefit from the insights gained while solving other aspects of the cosmology problem.

Organizing principle

Every successful venture into unknown territory requires relying on the appropriate beacons and guidelines and this is particularly true when the voyage takes you to the boundaries of traditional certainties and brings you to question some essential aspects of what had previously appeared to constitute a fixed background for scientific exploration. I would therefore like to briefly describe what was the essential principle that guided me on developing the revision of classical and semi-classical theories that is described in this report. It must first of all be understood that this principle was not given as a precondition imposed on any vision of the world, but actually developed alongside improvements in my and other people's knowledge and understanding of that portion of physical reality we actually experience and through the possibility that this probing allowed of inferring the regularities present in an even larger and more encompassing domain of the same reality.

My awareness of the importance of this principle developed mostly in conjunction with the elaboration of a more consistent appreciation of the requirements imposed by the classical theory of gravitation. Indeed, it is while tackling the problem of negative energy that I realized the importance (emphasized by others in a different context) of a relational view of the physical attributes of objects and that I understood the real significance of the requirement of relativistic invariance. This allowed me to perceive the true meaning of Einstein's insistence that the objects of physics must be conceived of only in relation to the spacetime structure to which they belong, because indeed I saw that the metric properties of space and time must be understood as dependent on the sign of energy of an object (as will be explained later), which is in contrast with traditional expectations. Thus, if a determination of the relationships between physical objects in different spatial locations or states of motion is possible only when we determine the common spacetime structure shared by those objects, then the fact that the spacetime structure itself is dependent on the nature of the objects means that the relationships

between them are dependent on their nature and in particular their energy signs. It therefore appeared to me that it is not only the position and state of motion of objects which require a relational viewpoint, but that any physical quantity must always be defined or characterized in relation only to similar quantities of other objects in the same universe (the physical attributes of a system enable to characterize it merely in relation to the similar attributes of other systems and those relationships are determined through the use of reference systems).

When I tried to understand what could logically impose such a requirement I slowly came to realize that it is the very fact that it would be meaningless to relate some physical quantity, in order to define its value, to some reference point not part of the same physical universe. Indeed, in the absence of a well-defined continuous network of causal relationships that would extend to those immaterial reference systems there can be no meaningful definition of the physical quantity involved, because physical relationships are material relationships and an object cannot be put into relation with something that is not part of the same causally related ensemble (the universe) to which it belongs. This requirement of a relational definition of physical quantities will have enormously important consequences on many aspects of the developments to be discussed in the following chapters.

Chapter 1

Negative Energy

1.1 The negative energy hypothesis

Regarding the question of negative energy, the current situation has much in common with that in which we were at the turn of the previous century with regard to the quantization hypothesis. There was in effect some reluctance initially to recognize the validity of the original suggestion by Max Planck that energy is quantized despite the fact that this proposal would have solved the problem of black body radiation. The trouble was of course that recognizing the validity of the quantization hypothesis would have required abandoning classical physics. There is a similar dilemma with negative energy today because, as I will show, this hypothesis has the potential to solve many important problems facing theoretical physics, but those benefits come at a price which may at first appear to be too high. Indeed, the introduction of negative energy matter as a concept somewhat distinct from that which is currently favored (which I believe is required in order to allow it to be consistent from a basic theoretical viewpoint) seems to imply that general relativity has to be abandoned. But rejecting a theory so well established and so beautifully simple as general relativity is not something that most people would do without very good motives. Yet, if the current assumptions concerning the rules governing negative energy matter (if it was to actually exist) may appear to better agree with relativity, they actually contradict some of the basic principles on which this theory is founded, therefore making it just as untenable. We must then either abandon the idea that negative energy matter can exist, or else provide a better interpretation of negative

energy states which may force a reinterpretation of relativity theory itself. But I will show that the conclusion that the latter alternative is the only viable one is not necessarily as dramatic in terms of its consequences as may seem, because what is required in this context is mainly a reinterpretation of the equivalence principle and not a rejection of the whole mathematical framework of relativity theory.

There is however an additional problem for the negative energy hypothesis which is that there appears to be no observational evidence for matter in such a state. But here also there is an analogy which should teach us a lesson. This is the case of the neutrino as a massive particle. For a long time when I was reading physics papers or any book on the subject of particle physics I could see that it was nearly *always* assumed, more or less implicitly, that the neutrino is massless as if this was a fact, while actually there was absolutely no evidence that this is actually the case and it was merely the difficulty to prove that the hypothesis is wrong that justified that everyone just assumed that the neutrino is massless. But just as for the idea that negative energy matter does not exist, I thought that it was incorrect to simply assume that the neutrino is massless when this could not yet be considered a fact. Thus, I always kept an open mind about those issues, because I saw that there were strong arguments (usually not recognized) for rejecting those commonly held assumptions and in the case of the neutrino at least it appears that this position was justified. In fact, I will later explain that there are very good reasons to expect that it should not be easy to confirm the existence of negative energy matter, because, as I have come to understand, it is not even directly observable, just as the more common, hypothetical dark matter. Thus, if I'm right, the implicit assumption that negative energy is forbidden would be just one of those 'reasonable' assumptions which we should be careful not taking too seriously.

The problem of negative energy has another parallel in a distinct but not entirely unrelated problem which is that of the origin of the arrow of time. Indeed, it was suggested by some eminent figures that the problem of irreversibility could be solved by integrating some fundamental element of irreversibility into the formalism of even the most elementary physical theories. This would seem to be justified by the fact that the problem of time asymmetry has been known to exist for a long time and no acceptable solution to it that would be based on boundary conditions imposed on otherwise time-symmetric evolution has ever been found. But again I think that the difficulty to prove a hypothesis (that time asymmetry can arise from time-symmetric

physical laws) should not be taken as evidence that what may perhaps be its only alternative (that time asymmetry is fundamental) is right. In the case of negative energy, we are also in a situation where we have built into the very formalism of our most fundamental theory of matter (which currently is quantum field theory) the apparently necessary, but clearly unjustified (from a theoretical viewpoint) hypothesis that only positive frequencies (associated with positive energies) are allowed to propagate in the future (the constraint on negative frequencies being merely that they must propagate toward the past).

However, I think that the fact that this artificial restriction appears to be valid does not imply that positive frequencies cannot propagate backward in time or that negative frequencies cannot propagate forward in time, but merely that if there exist two kinds of matter related by their opposite energy signs (the frequency signs relative to the direction of propagation in time) then, for some reason, they can only interact with matter of the same energy sign (I will eventually explain why such a limitation naturally occurs). This absence of interaction or interference (in the classical sense) is what really justifies that quantum field theory only deals with matter of one energy sign under most circumstances (when gravitation is not involved). But given that I'm suggesting that energy sign is a relatively defined physical property, so that there is no absolute (non-relational) distinction between positive and negative energy matter, then it must in effect be concluded that there cannot exist a constraint that would impose that negative energy matter and only matter with such an energy sign does not exist if positive energy matter itself is allowed to exist, as required, because it is not even possible to identify the distinguishing property specific to negative energy matter that would justify that its existence be ruled out. Thus, I'm allowed to conclude that any attempt at getting rid of the apparently intractable problem of negative energy states by simply imposing a constraint to be applied on the formalism itself is misguided and unnecessary, because, indeed, once an appropriate understanding of the true nature of negative energy matter is available it becomes apparent that a restriction on allowed frequencies is no longer necessary. In fact I believe that the same can be said of the problem of irreversibility, because in chapter 3 I will show that the thermodynamic arrow of time is not an intrinsic feature of fundamental physical laws, but instead originates from an unavoidable constraint that applies on the boundary conditions at the Big Bang.

In the context where we must recognize that there is no motive to reject

the possibility that negative energy matter may be present in our universe it becomes apparent that one often mentioned argument that must definitely be rejected concerning the nature of the gravitational interaction is the idea that the strength of gravitation on the largest scales is a consequence of the 'fact' that this interaction is always attractive. This is a conclusion which is usually assumed to follow from the observation that there does not exist negative gravitational charges (negative energy matter is assumed not to exist). Yet, what actually explains the fact that gravitation is a dominant force on larger scales (in addition to its long range property) is not the absence of matter in a negative energy state, but the simple fact that gravity is attractive between objects with the same positive gravitational charge, that is, between objects with a positive sign of energy. Thus, if gravitation dominates over electrical forces on astronomical scales it is really a consequence of the fact that while identical electric charges tend to disperse under mutual electrostatic repulsion, positive energies have a tendency to coalesce and to accumulate under mutual gravitational attraction and the fact that electromagnetism is already known to have both positive and negative charges has nothing to do with the fact that those charges do not so readily accumulate, because even if there were only positive electric charges they would not cluster, because identical electric charges mutually repel one another and the possibility for such opposite charges to cancel out actually facilitates an *accumulation* of those charges, but only in neutral configurations and under the influence of gravitation.

It must therefore be understood that there is no requirement for gravitation to always be attractive merely on the basis of the fact that its existence can be felt despite its extreme weakness, as is sometimes suggested. Indeed, if it was found that there actually exist negative energy particles, the possibility for energy to cancel out would not necessarily prevent the accumulation of matter with one or another energy sign, because negative energy matter may also be gravitationally attracted to itself (despite what is usually assumed) and could therefore also be subject to accumulation. To summarize, what makes electrical forces negligible on the large scale is the fact that identical electric charges do not attract one another and therefore do not accumulate as may identical gravitational charges. Instead electric charges of opposite signs are attracted to each other and immediately cancel out, therefore preventing further accumulation, at least under the influence of electric forces. But this does not mean that gravitation would be submitted to the same fate if negative energy particles were found to exist, because it may well be

the case that gravitational charges with the same sign always attract one another given that this is already known to be true for positive energy matter and this would not even forbid opposite energy bodies from gravitationally repelling one another. The frequently encountered remark that gravitation is attractive for all particles should therefore be understood to mean only that it is attractive for all currently known forms of matter.

Thus, again, the observation of large accumulations of positive energy matter is not an argument against the existence of negative energy matter. But it is also true that the apparent absence of large accumulations of *negative* energy matter would not necessarily mean that such matter does not exist, even if we were to assume that this matter gravitationally attracts matter of the same kind. Indeed, it may turn out that this matter is dark and given that it may also be repelled by positive energy matter (even if this is not what we usually assume) then we might be justified to expect that it should be located mainly in regions of the universe where the density of positive energy matter is the lowest. Therefore, negative energy matter would be virtually absent from regions where positive energy matter is more abundant, like that in which we are located, and this would explain that we have never noticed its existence. I will explain later why the assumptions discussed here concerning the nature of negative energy matter should in effect be those which are retained, thus confirming the validity of the above explanations as to why it is that negative energy matter appears to be absent from our universe. It will then be clear that theoretically it is to be expected that if negative energy matter exists it should have the properties which are responsible for our very ignorance of its existence.

I think that what must be recognized above all is that the commonly held view that the occurrence of negative energy in a theory is necessarily always indicative of a problem is not rationally motivated and that it is not true that all traces of negative energy *must* be eradicated at all costs whenever they are encountered. Dirac, at least, understood that the prediction of negative energy states could not be ignored and tried to provide an explanation for the absence of transitions to such states [1]. His solution, based on the idea that negative energy states are already all occupied, was not satisfactory, but at least he did not simply reject the possibility that negative energy matter might have to be considered real. There is no motive to argue, as people often do, that negative energy is totally unacceptable, other than the difficulty to find an appropriate interpretation that would be compatible

with empirical facts for this logically unavoidable counterpart to positive energy. In the absence of a theoretical justification for the absence of negative energy matter I think that the only appropriate approach would be to seek to find out why it is that we never observe matter in such states, rather than try to build that assumption into a then necessarily incomplete theory of quantum fields. In this particular sense it is significant that the prediction of antiparticles was a by-product of Dirac's original interpretation of negative energy states, because this contributed to the belief that the discovery of antiparticles constitutes a definitive solution to the negative energy problem. But, given that Dirac's interpretation was later found to be inappropriate, I think that we need to recognize that in fact antiparticles can only be one particular aspect of a complete solution to the problem of negative energy, which therefore remains unsolved.

In any case it must be understood that even if we were to succeed in justifying that it should be imposed that there cannot be transitions from a positive energy state to a negative energy state, we would not have solved the problem of negative energy. This is because such a restriction would merely impose that no positive energy particle can turn into a negative energy particle (and vice versa maybe), but there would be nothing in that constraint to forbid a particle to already be in a negative energy state, in which case we would still need to provide a consistent description of the properties of matter in such a state and to justify that we do not observe those negative energy particles under most conditions. In fact, I will later provide arguments to the effect that just such a restriction on energy sign shifting transitions is to be expected to occur very naturally, even if negative energy matter must indeed be allowed to exist. Anyhow, the fact is that if there is no reason to assume that some restriction applying to energy sign reversal would forbid *positive* energy matter from existing then there cannot be more justification in assuming that such a restriction forbids negative energy matter from being present in the same way. I must insist again that there is no reason to assume that the concept of negative energy is problematic all by itself and that negative energy must be avoided systematically, because the only requirement, regarding negative energy states, may be that there cannot be transition to such states by a particle in a positive energy state and this only when the transition would be to a state of negative energy propagating forward in time. Such a requirement is necessary (although not entirely sufficient) to keep positive and negative energy matter virtually isolated at the quantum level, so that the experimental constraint of an absence

of interference from negative energy matter into the theoretical predictions involving positive energy matter can be satisfied.

I do understand of course that there are a number of issues associated with the possibility that matter may occupy negative energy states. Of particular concern would be the issue of 'vacuum decay' or the apparent problem that all positive energy particles should fall within a very short interval of time into the available negative energy states by releasing a compensating amount of positive energy radiation, if those states are not assumed to be forbidden. In fact, this problem would seem to affect negative energy matter itself, even if transitions to negative energy states by positive energy particles were found to be impossible. This is of course the difficulty that motivated Dirac's problematic proposal that those energy states should already be nearly completely filled so that no further decay should occur. But I will show in later portions of this chapter that this problem and also some others which may seem to arise in relation to the possibility for negative energy matter to exist in a stable form are merely a consequence of the inappropriateness of the current interpretation of the concept of negative energy. In fact, it will be shown that it is not even necessary to assume that negative energy states cannot be reached by matter in a positive energy state, because even matter already in a negative energy state cannot be assumed to fall to even 'lower' energy states.

I also recognize that the tentative interpretation of negative energy states that came to replace Dirac's solution does in effect provide some level of relief in that it at least allows to take into account those negative energy states that cannot be ignored as they actually interfere with processes involving ordinary matter. This is because we are indeed allowed to consider that antiparticles are negative energy particles propagating backward in time. But even under that particular interpretation, antiparticles can still be conceived as ordinary particles (submitted to normal gravitational interactions) from the forward time perspective relative to which their energy is positive and therefore they cannot be considered to provide an interpretation of negative energy states of the kind that would be truly significant from a physical viewpoint. Again, the exclusion of true negative energy states may appear to be justified from an observational viewpoint, but it still constitutes an arbitrary rule which would at least require an explanation, as there is no consistency principle behind it. It is therefore certainly amazing that so many otherwise well informed authors suggest that no negative energy, or negative mass particle can exist, as if this was an obvious and unavoidable conclusion. It must be

clear that I'm not complaining about this situation, I merely want it to be recognized for what it is, because I will take a different course and it should be understood that I'm not doing this without good motives or out of a fondness for hopeless, exotic or eccentric ideas.

I must therefore mention that I'm aware that the originators of the steady state theory of cosmology once also criticized (based on distinct motives) the traditional position according to which the existence of negative energy matter is forbidden. But if I do find this criticism to be valid and appropriate I do not, however, find suitable the whole concept of negative energy (which is actually very traditional) proposed by these authors, nor do I agree with the objectives they unsuccessfully (given the failure of steady state cosmology) sought to achieve by using this otherwise interesting idea. I think that the fact that the hypothesis of negative energy matter was historically associated with such failed theoretical models and was also developed into many different inconsistent formulations lacking any epistemological support is more than anything else responsible for the state of suspicion and confusion that currently surrounds the whole idea of negative energy matter. The objective I will try to achieve in this chapter will therefore be to clarify the situation regarding what should be expected regarding the properties of matter in a negative energy state and to demonstrate the validity of the concept itself in the context where it is properly conceived and justified.

1.2   The time-direction degree of freedom

What emerges from my re-examination of the assumptions behind our current understanding regarding the possibility that particles may occupy negative energy states, is that we must first recognize that for any elementary particle there exists a fundamental degree of freedom related to the direction of propagation in time of its charges, including the gravitational charge, that is to say, including energy. The existence of such a degree of freedom means that a positive charge can in effect be positive either in relation to the positive direction of time, if such a charge propagates in the positive direction of time, or in relation to the negative direction of time, if the same positive charge propagates in the negative direction of time. But the particles so characterized would be physically different from one another. It is not possible therefore to completely specify the physical properties of a particle at a given instant by simply providing the sign of its charges independent

from their direction of propagation in time. But given that a particle can actually be identified by the charges (including energy) it carries (it has no other physical properties except for its momentum, position, and spin at a given time) this means that the apparent nature of a particle may depend on whether it propagates its charges in the positive or the negative direction of time, that is, it may depend on whether it is itself propagating forward or backward in time[1]. The physical attributes of a particle can only be unambiguously defined in relation to the direction of time in which this particle propagates and this is true also for energy.

This is what the insights gained by considering the consequences of the relativity of simultaneity for the quantum description of particle interactions should be understood to imply. Indeed, it is the fact that some processes involving the exchange of a virtual particle of interaction cannot be assigned a unique definite order of occurrence in time that renders the notion of particles propagating backward in time unavoidable. This is because the emission and absorption events of such an exchange process are spacelike separated so that their order of occurrence in time is dependent on the state of motion of the observer. Thus, what is viewed by one observer as the emission of some particle carrying a negative charge, can be seen by another observer as the absorption of a similar particle carrying a positive charge, which certainly requires the sign of charge to be dependent on the perceived direction of propagation in time. Given the undeniable validity of this viewpoint, the only argument that could still allow one to reject the reality of a degree of freedom associated with the direction of propagation in time would be one based on the second law of thermodynamics and the apparent impossibility for a macroscopic body to 'travel' backward in time. It appears, however, that this argument is not valid, because the thermodynamic constraint only applies to the flow of information as it occurs through the formation of records

---

[1]I'm here considering a particle in a semi-classical way, as if we could always associate with it a definite position and momentum, even though it is clear that actual knowledge of those conjugate attributes cannot be obtained at the same time. This idealization simply allows to gain insight into what would be the properties of an elementary particle if it could be observed at the energy scale of an actual macroscopic body, while still carrying a mere unit of its other charges. We may alternatively consider a real macroscopic body and assume that it has physical properties that evolve in a perfectly coordinated fashion, with all its charges necessarily propagating in the same direction of time at all times (therefore acting as one 'macroscopic' charge), but such a viewpoint is actually even less realistic than the former idealization (for reasons that will appear more clearly later on) and would change nothing to the following conclusions.

and in no way forbids individual particles from propagating backward in time as long as they are not involved in processes which (collectively) would allow information to be transferred from the future to the past. It is therefore merely this limitation on the flow of information that explains the fact that our experience of reality has made us suspicious of the possibility that objects themselves (or particles) can propagate backward in time and not the actual impossibility of such an occurrence.

In such a context the possibility to distinguish the sign of a charge, including energy, would depend on the possibility to determine the direction of propagation in time of this charge. Thus, even independently from the argument based on the relativity of simultaneity, we may consider that the sign of charges and in particular the sign of energy is defined only in relation to the state of motion of the particle carrying those charges, where 'motion' is here relative to time instead of space. But if we may also assume that the attribution of a direction of propagation in time is merely a matter of convention, because all that can be asserted is whether any two particles are propagating in the same direction of time or in opposite directions, as I will suggest later, then it would appear that the sign of energy itself would become a relative notion dependent on which direction of time is chosen as that in which a given particle propagates. In this particular sense we would have to recognize that associated with the relativity of 'motion' in time there is also a relativity of the sign of energy.

Acknowledgement that the sign of energy is a relative property actually allows one to reject the validity of the constraint usually imposed that all energy must be positive, because it means that even what appears to be positive energy according to one particular convention for the direction of propagation in time is actually negative energy according to an alternative choice for the same time-direction parameter. The possibility for particles to propagate backward in time, which is made unavoidable by the fact that backward in time motion is actually required under a consistent understanding of the constraints imposed by a relativistic treatment of quantum processes, as mentioned above, therefore actually implies that negative energies must also be allowed in physical theory, because even what we usually describe as a positive energy particle could be redefined as a negative energy particle if we were to also assume as a matter of convention that the direction of propagation in time of the particle is opposite that which is usually (more or less implicitly) assumed. Negative energies must be considered to be possible states of matter even if only for particles propagating in the backward

direction of time. This dependence of energy sign on the assumed direction of propagation in time is what allows antiparticles to actually be described as particles propagating backward in time with negative energies and unchanged non-gravitational charges as Feynman once suggested [2], even if we are also allowed to consider those particles as positive energy particles with reversed non-gravitational charges propagating in the usual forward in time direction.

What is essential to understand here is the dependence of the value of any charge, including energy, on the direction of time in which this charge is assumed to be propagating. Thus, simply saying that a particle has positive electrical charge or positive energy doesn't make sense. We must also always specify the direction of propagation of this energy with respect to the time parameter. What appears to be a positive charge or a positive energy relative to the positive direction of time would be a negative charge or a negative energy relative to the negative direction of time. Thus, all those energy signs are merely established on the basis of practical conventions and can never be asserted in an absolute fashion. It must be recognized, however, that if the energy of an electron is by convention considered positive relative to the future direction of time in which it is, again by convention, assumed to propagate, then the energy of an anti-electron must *necessarily* be considered negative relative to the past direction of time in which it must, under the same convention, be assumed to propagate. It is merely because we ignore the requirement to describe the positron as propagating backward in time that we can attribute to it a positive energy. As a consequence, it would seem that even on the basis of current observations we would not be allowed to assume that particles are forbidden from occupying properly defined negative energy states.

Yet despite the unavoidable character of this conclusion and even in the face of the enormous simplification of our world view that is made possible by the hypothesis of the existence of a fundamental degree of freedom related to time direction, it is still often suggested that the interpretation of antiparticles as particles propagating backward in time with negative energy is merely a mathematical artifact and corresponds to nothing real. But I think that this attitude is similar to that of nineteenth century philosophers and scientists rejecting the hypothesis of the existence of atoms, even in face of the overwhelming evidence in favor of this concept, supposedly because the atoms could not be seen directly, but actually because of an unjustified prejudice in favor of a continuous, macroscopic description of matter. Given

the above discussion concerning the relative nature of energy sign, I think that it is clear that there is no basis for assuming, as is often done, that the negative energy of antiparticles as particles propagating backward in time is not real and that those particles are merely 'ordinary' particles which happen to be carrying opposite non-gravitational charges. If we are allowed to describe antiparticles as particles propagating backward in time, then we must recognize the existence of negative energy states.

It must, in this context, be understood that the commonly met suggestion that all physical properties are simply reversed for an antiparticle (by comparison with those of the associated particle) is wrong, because the signs of all physical quantities are dependent on the direction of propagation in time and we would at least have to specify with respect to which direction of time the various quantities are to be assumed reversed. Indeed, even from the viewpoint where antiparticles are assumed to propagate in the same direction of time as do regular particles we would have to admit that energy is not reversed for an antiparticle, otherwise a pair annihilation process should release few or even no energy in the form of radiation, contrarily to what is routinely observed. Also, if we do consider instead the viewpoint of an antiparticle's true (when ordinary particles are assumed to propagate forward in time) direction of propagation in time, then energy would indeed be reversed as I already mentioned, but all non-gravitational charges far from being reversed would have to be considered rigorously unchanged given that from the forward in time viewpoint they actually appear to be reversed while from my perspective the sign of charge is a relative notion dependent on the assumption that is made regarding the direction of propagation in time of a particle.

Thus, what appears to be a positively charged particle in relation to another particle propagating forward in time would actually appear to be a negatively charged particle in relation to yet another particle propagating backward in time and the same would be true of energy sign. Those relative alterations of the sign of charges occurring as a consequence of a reversal of time are manifested merely in the fact that what is found to be a repulsive non-gravitational interaction between two identical particles propagating in the same direction of time, would upon a reversal of the direction of propagation in time of one of the particles become an attractive interaction, or vice versa, as a result of the equivalent reversal of the sign of charge that occurs when a particle reverses its direction of propagation in time without actually reversing its charge. This is an unavoidable consequence of the fact that the

departure of a positively charged particle from a region of space would from a reversed time viewpoint necessarily appear as the arrival of a particle of opposite (negative) charge, therefore implying that there is a relationship between the relative direction of propagation in time and the relative sign of any conserved physical quantity. We do not even have to know what an electric charge is or what energy is from an exact mathematical viewpoint to draw that conclusion. The reversal of charges associated with a reversal of time simply illustrates the subtlety of the relational definition of the sign of conserved (time-invariant) physical quantities in the context where there is a fundamental degree of freedom associated with time direction.

It must be remarked that in the context where there is in effect a dependence of the sign of charges on the direction of propagation in time it follows that there no longer needs to be a mystery regarding why all charges come in two varieties, each having the exact same magnitude, but a polarity opposite that of the other. This is because even if there were only, say, positive electrical charges, the fact that particles are free to propagate either forward or backward in time (under appropriate conditions) means that from a practical viewpoint there would still occur phenomena involving negatively charged, but otherwise identical particles and it would not be possible to say whether it is the positive or the negative charges which constitute the 'true' charges. In such a context it seems possible that the requirement imposed by grand unified theories that the sum of charges of all elementary particles cancel out, so that the overall symmetry is preserved in the context where it is not spontaneously broken, could ultimately be understood to be made possible (if the current elementary particles are actually composed of more fundamental building blocks) by the relativity of the sign of charges with respect to the direction of time, which not only allows, but actually requires the existence of opposite charges. What I'm now suggesting is that we would in fact be justified to consider that the same requirement also applies to energy, which would therefore come in two varieties with opposite signs, not only for particles propagating in opposite directions of time, but even relative to the conventional forward direction of time.

In any case it should be clear that it is no longer possible to consider the sign of charges, including that of energy, independently from their direction of propagation in time. The traditional viewpoint according to which it seems possible to define charge without reference to some direction of time is valid merely because we implicitly always consider the sign of charge with respect to the positive direction of time (conventionally assumed to be the

future). The positive definite value of energy under all circumstances is thus an artifact of this implicit choice of the positive direction of time as the direction relative to which energy is measured. It is true though that if it was not for the non-gravitational charges carried by a particle it would in effect be impossible to distinguish between the case of a positive energy propagating forward in time and that of a negative energy propagating backward in time, just as it would be impossible to distinguish between the case of a negative energy propagating forward in time and that of a positive energy propagating backward in time. But there is no reason to assume that there would be no distinction between positive and negative energies propagating in the same direction of time and therefore the truly significant measure concerning energy is the sign of action, which is obtained by multiplying the sign of energy by the sign of time intervals. If the hypothesis that energy must necessarily be positive has always appeared valid it is merely as a consequence of the fact that we always measure energy relative to the positive or forward direction of time and for all known particles action remains positive. As I suggested above, however, this does not mean that energy really is always positive, but merely that action, or the sign of energy relative to the sign of time intervals, is in effect always positive for all currently known particles, independently from the true sign of energy of those particles.

What I would like to suggest, ultimately, is that in fact it is not only the sign of energy that is to be viewed as a relative quantity, but that the sign of action itself is purely relative, in the sense that there could never exist a generally agreed absolutely defined positive or negative value for the sign of action of a particle. In this context not only would the sign of energy be dependent on the direction of time in which a particle is assumed to propagate, but the sign of action would itself depend on the choice of what direction in time is to be that in which what are assumed to be positive energy particles propagate, or what is the sign of energy of those particles which are considered to propagate forward in time. Here all that matters is that once you define one particle as having positive action, because you assume that it is this particle that propagates positive energy forward in time, then the particles that you must assume to be carrying negative energies forward in time or positive energies backward in time as a consequence of this *choice* are those which will have negative action. But it must be clear that you are always free to describe the first particle as propagating negative energy forward in time and therefore as having negative action, as all by itself this choice is arbitrary, but in this case the other particles would then

necessarily have to be assumed to carry positive action instead of negative action, because their *relationships* of time directionality and energy sign with the first particle (the difference or the identity of the signs of time intervals and energy) would remain unchanged.

It must also be remarked that the fact that what we would currently define as negative action particles are related to ordinary matter through a simple convention regarding the direction of propagation in time means that the motive for rejecting the possibility that negative action matter may actually exist is no stronger than that which would consists in arguing that ordinary matter itself is not allowed to exist. There is absolutely no rational motive for rejecting the viewpoint described here and many reasons to recognize its validity. In any case the fact that the sign of action is a purely relative concept which can vary as a consequence of assumptions regarding the direction of propagation in time means that if the direction of a local gravitational field depends on the sign of action of its source then it should itself vary as a function of the assumptions made concerning the direction of propagation in time of the objects submitted to it (which determine their own action signs in relation to that of the source) and therefore the gravitational field must itself be considered a relative concept dependent on the conventions used by an observer.

Regarding the relation between the sign of charges in general and the direction of propagation in time it must be noted that energy actually distinguishes itself from non-gravitational charges by the fact that it is naturally reversed when a particle reverses its direction of propagation in time. Indeed, in the context where a particle-antiparticle annihilation process must be considered as an event during which a particle bifurcates in time to begin propagating the same non-gravitational charges backward in time (which would effect the same kind of change as reversing the charges and keeping the direction of propagation in time unchanged), it must be assumed that the energy of the particle is reversed along with the direction of time intervals when the bifurcation occurs given that the particle now propagates backward in time while its energy remains positive from the conventional forward in time viewpoint. In fact we have no choice but to consider that only non-gravitational charges are left unchanged (relative to the true direction of propagation in time) when the particle begins propagating backward in time during what appears to be a particle-antiparticle annihilation process, because energy is always released by such a process and if the sign of en-

ergy had remained unchanged along with that of non-gravitational charges when the direction of propagation in time of the particle reversed, then an antiparticle's energy would be opposite that of its particle with respect to the forward direction of time and therefore the annihilation of such a pair could occur without any energy at all being released, as I previously mentioned. Thus, energy must actually reverse along the 'true' direction of propagation in time of a particle, when the particle reverses its direction of propagation in time during a pair annihilation process, just like momentum naturally reverses when a particle changes its direction of motion in space. The negative energy of an antiparticle simply propagates backward in time so that relative to the positive or forward direction of time it is left unchanged and from a mathematical viewpoint this interpretation fully agrees with the traditional description.

If this relational interpretation of the energy signs of particles involved in pair annihilation processes is valid then, based on the fact that we also have many reasons to believe that the gravitational properties of antiparticles are the same as those of particles, I can deduce that from a gravitational viewpoint the sign of energy is physically significant merely in relation to the direction in which a particle with that sign of energy is propagating in time. In other words, to produce an anomalous gravitational field, or to respond anomalously to a gravitational field, a particle would have to propagate its negative energy forward in time rather than backward, as does an ordinary antiparticle. This is a simple, but very significant result whose consequences will be developed in the following sections. What must be understood is the fundamental character of the degree of freedom associated with time direction, which in a general relativistic context simply embodies the sum of all relationships of time directionality between a given particle and all the other particles in the universe. This physical property must be considered distinct from any property of time directionality which is merely statistically significant and which is associated with the flow of information, as that which characterizes the irreversible processes obeying the second law of thermodynamics.

Concerning the gravitational properties of antimatter, it appears that it is actually unnecessary to appeal to any independent constraint like the equivalence principle (which seems to require all matter to have the same acceleration in a gravitational field) to justify that antimatter should not 'fall' up in the gravitational field of a positive energy planet like the Earth, as was often

proposed before experiments began to rule out such a possibility. Indeed, any of the arguments traditionally provided to rule out the possibility of an anomalous gravitational behavior of antimatter become unnecessary once it is understood that it is actually only matter propagating its negative energy forward in time that could experience gravitation distinctively from normal matter, while it is already known that if negative energy is to be associated with antiparticles then this energy would in fact propagate backward in time. There is thus a very good reason to assume that antimatter falls down in the gravitational field of the Earth, but this is not an argument that we could use to rule out the possibility that some matter that would not be antimatter could perhaps be subject to anomalous gravitational interaction with ordinary matter, because there is no a priori motive for assuming that there cannot exist particles propagating negative energy forward in time. In fact, I will later explain that even the general argument against anomalously gravitating matter based on the necessary application of the equivalence principle is not really unavoidable, because it is possible to better define this principle in a way that allows for the existence of anomalously gravitating matter of the appropriate type, while retaining the general framework of relativity theory which can accommodate such a generalization.

In any case it must be recognized that all those properties of fundamental time directionality discussed above are a reflection of the fact that the sign of charges (including energy) is not only defined in relation to the direction of propagation in time of the particle carrying those charges, but is actually determined completely arbitrarily as being merely significant in relation to the similar physical properties of other particles. From a relational viewpoint it would be incorrect to assume that the direction of propagation in time of a given type of particle, carrying a unit of electric charge with a given, arbitrarily assigned positive or negative sign, is definitely the future direction, say, while the direction of propagation of the antiparticle of the same type is definitely the past, or even that there exists an absolutely defined character of being an antiparticle by opposition to being a particle. The only physical property that can be objectively defined without referring to quantitative attributes of objects that are not part of our universe is the relative direction of propagation in time of two particles. Two particles with the same type of charge may be both propagating in the same direction of time or they may be propagating in opposite directions of time and this is all we can ascertain through physical means.

What must be understood is that while the relationship between the

direction of propagation in time and the sign of a given charge, including energy, is a matter of coordinative definition (a definition that must be applied similarly to all processes in the whole universe on the basis of their relationships to one particular process for which an arbitrary choice of properties is assumed), once such a definition is applied the difference between the sign of time intervals and the sign of charges is an objective physical property that is not dependent on a particular viewpoint. But it is not just the relationship between the sign of charge and the direction of propagation in time of a particle which can be given clear meaning through the use of a coordinative definition, because once we define one kind of particle as actually propagating a positive charge forward in time then it should also be possible to differentiate such a particle from an otherwise identical particle propagating a negative charge in the opposite direction of time.

It must be clear, therefore, that once we assume an ordinary electron to be propagating its negative charge forward in time it is not possible to consider another *ordinary* electron as perhaps propagating backward in time while carrying a positive electric charge in this direction of time (so that the electron would still appear to be propagating a negative charge relative to the forward direction of time). Indeed, if a certain condition of continuity of the flow of time on which I will elaborate in section 2.10 is assumed to apply, such a backward in time propagating ordinary electron could only annihilate with an anti-electron which would be propagating the same positive charge forward in time (instead of propagating a negative charge backward in time). But this would actually mean that certain positrons cannot annihilate with certain electrons while no constraint of this kind is observed to apply, as all known electrons have the same unique probability of annihilating with any positron. Thus, if a constraint of continuity of the flow of time along an elementary particle world-line does indeed apply, an ordinary electron must be assumed to propagate in one and only one direction of time while its antimatter counterpart must similarly be assumed to always be propagating in the opposite direction of time. Perhaps that this restriction is a consequence of the fact that there actually exists only one electron or that all electrons are 'the same particle' propagating forward and backward in spacetime, as John Wheeler once argued, but the condition of continuity of the flow of time does not specifically require the validity of this hypothesis.

On the basis of those considerations and given the previously reached conclusion that only the sign of energy with respect to a given direction of time has physical significance, it must in effect be recognized that only a

particle propagating either negative energy forward in time or positive energy backward in time (in the context where ordinary matter is considered to propagate positive energy forward in time) could potentially respond in an anomalous way to the gravitational interaction. What is important to know about such a particle, which we may call a negative action particle[2] to distinguish it from a particle merely propagating negative energy backward in time like an antiparticle, is that the preceding considerations regarding the relational definition of physical quantities would also mean that the particle cannot possibly be considered to have physical properties that would qualify it as responding to the gravitational field of a positive action body in an anomalous fashion that would not also be shared by an ordinary matter particle (propagating positive energy forward in time) submitted to the gravitational field of a negative action body. This must be considered an unavoidable conclusion in the context where one can physically distinguish only a difference or an equality in the signs of action of any two particles and cannot attribute objective meaning to the sign of action itself. That does not mean that there would actually be no anomalous response, only that in a configuration where all 'anomalously' gravitating matter is replaced by ordinary matter and all ordinary matter is replaced by anomalously gravitating matter we should observe no difference (for the most part). Thus, a particle defined as having negative energy relative to the positive direction of time and which would be located in the gravitational field of a planet having opposite energy relative to the positive direction of time should behave in the same way as a positive energy particle in the gravitational field of a negative energy planet and similarly for any combination of energy signs of particle and planet, because only the relative difference in forward propagated energy signs can be considered significant. Given the preceding discussion this should be crystal clear. But that is not what is usually assumed to occur by people discussing negative energy or making quantitative predictions involving matter in such an energy state.

What is usually assumed is that a positive energy or positive mass body would attract all bodies, regardless of whether those bodies have positive or negative energy or mass, while a negative mass body would repel all bodies, again regardless of whether those bodies have positive or negative mass. It

---

[2]Despite the ambiguity I still use the term 'negative energy' in place of 'negative action' to identify such anomalously gravitating matter when the context clearly indicates that I mean negative energy propagating forward in time or equivalently positive energy propagating backward in time.

is currently believed that this is the consequence of taking inertial mass to be reversed along with gravitational mass, as would appear to be required by the equivalence principle. It must be clear however that those are not results which are 'derived' from relativity theory as is sometimes suggested, but merely the consequence of a choice that is implicitly made regarding what properties should be associated with negative inertial mass while trying to be as accommodating as possible with the traditional conception of the principle of equivalence. But if I find it appropriate and indeed necessary to consider, as most people do, that inertial mass is reversed along with gravitational mass when we are considering an object with negative energy (so that the equivalence principle can be observed to apply), I cannot agree with the conclusion that is usually drawn from such an assumption. Indeed, for the response of various masses to the presence of a negative mass to be in line with common expectations, it must be possible to determine the sign of mass or the sign of action of particles in an absolute non-relational manner, because we are assigning the attractive or repulsive character of the gravitational field in precisely such an absolute manner (the field is either repulsive for everything or attractive for everything) which I believe could never be justified.

I think that it cannot be assumed that a negative mass is repulsive in an absolute invariant way, because it would not be possible to tell relative to what reference point the distinctiveness of this character is defined given that positive mass cannot be used as a reference if it is itself absolutely defined (not merely in relation to the opposite negative masses). I will explain in a later section of this chapter why it is that the assumption that a negative inertial mass is associated with a reversal of the sign of action, far from having the undesirable consequence of allowing absolutely defined physical properties into physical theory (if there could ever be such a theory) actually gives rise to a description of the gravitational interaction between positive and negative mass bodies that is in perfect agreement with the requirement of relational definition of the sign of mass or energy (once the inertial properties of negative mass matter are well understood). All that would then remain to understand is how the equivalence principle can still be satisfied by such a description. For that purpose I will provide arguments to the effect that a simple reconsideration of the true significance of the principle of equivalence, and a better understanding of its motivation in the principle of relativity of accelerated motion, allows its foundations to be preserved while enabling the more consistent relational viewpoint on the sign of mass to be retained and

to actually be integrated into the core mathematical framework of relativity theory by introducing a slight modification to this classical theory of gravitation that is actually a simple generalization of it. In order to further justify this approach, I will first try to identify what should be the true properties of negative action matter and why we should not expect such matter to behave in ways that would make it undesirable not only from the viewpoint of the requirement of a relational description of physical quantities, but with respect to other constraints and other physical principles which we can be confident must also be obeyed.

1.3 Our current understanding

Before addressing the question of how a negative energy particle would actually behave we may first want to explore what the current situation is regarding the notion, or indeed the problem of negative energy. For this purpose, it should first of all be noted that for many reasons no one seems to like the idea that there could exist negative energy particles. Thus, it is no surprise that one of the most basic and often implicit assumption that enters our description of physical reality is that energy must always be positive. There are many different formulations of that requirement which impose various degrees of conformity to the hypothesis that matter cannot find itself in a state that would be observed as having negative energy. In its least restrictive form this condition is called the weak energy condition and merely constitute a statement about the positivity of the components of the stress-energy tensor (the most general representation of the energy content of matter). More constraining conditions have also been proposed, among which is the appropriately named strong energy condition which if obeyed under all circumstances would mean that gravity must always be attractive (between all forms of matter which would then be allowed to exist). Those conditions are used as rigorously defined hypotheses in various theorems dealing with the behavior of matter under the influence of the gravitational interaction.

The problem is that it was found at some point that configurations involving negative energy densities are actually allowed to occur in quantum field theory [3]. This does not mean that negative energy particles are explicitly allowed by current theories, but merely that unlike what we would expect from a classical viewpoint where the vacuum is described as a total absence of matter, quantum field theory allows for the local density of en-

ergy to not always be positive definite, even in a context where only positive energy matter is present. A well-known experiment illustrates the kind of phenomena involved. It requires placing two parallel mirrors a very small distance apart in a vacuum so as to forbid some states, which would normally exist in the vacuum, from being present in the space between the mirrors, as a consequence of the incompatibility of their characteristic wavelengths with the spatial constraints imposed by the presence of the mirrors. The predicted result, which is actually observed, is that there should arise a small pressure pulling the mirrors together as a consequence of the comparatively larger pressure exerted from the outside, which is actually caused by a decrease in pressure from between the mirrors that can be attributed to the restriction imposed on which virtual particles can be present in this volume. This is of course the phenomenon known as the Casimir effect [4]. It is clear though that we are not directly measuring a negative energy density in such an experiment, but merely the indirect effects of an absence of some positive contribution to vacuum energy, which is then assumed to imply that the energy density is negative in the small volume between the mirrors. But even that kind of manifestation of negative energy is assumed to be so serious a problem by some theorists that they suggested that the description of the vacuum as involving virtual particles coming in and out of existence is actually only a mathematical trick and does not reflect what is really going on in the absence of 'real' matter.

However, this aversion for whatever is negative of energy is not shared by all authors and some more open-minded specialists have tried to address the issue of negative energies as they occur in quantum field theory and in so doing gained some significant insights into what exactly is allowed by a quantized description of the vacuum. A modified version of the weak energy condition was thus proposed that allows to take into account the fluctuations of energy which arise in the quantum realm. This condition, which is appropriately called the averaged weak energy condition, involves only quantum expectation values of the stress-energy tensor averaged over some period of time during which the observations are assumed to occur, rather than idealized measurements at a spacetime point. A feature of the constraint provided by this condition is that it allows for the presence of large negative energies over relatively large regions of space if there is a compensation by the presence of a larger amount of positive energy during the time period over which the observations are made. It was indeed found out [5, 6, 7, 8] that quantum field theory places strong limits on the values

of negative energy density that can be observed over finite periods of time under various conditions. What emerges from those developments is that there appears to be a constraint on the magnitude of negative energy that can be observed and it indicates that negative energy can be merely as large as the time interval during which it is measured is short. I believe that this is indicative of the fact that while negative energy states cannot be ruled out as strictly forbidden, they should also clearly not be expected to materialize in stable form in the context where we are dealing with ordinary matter configurations for which the particles are already predominantly in positive energy states.

A similar limitation can also be observed to restrain another form of negative energy that occurs in the presence of an attractive force field, even in a classical context. Indeed, the energy contained in the force field between two particles submitted to an attractive interaction must be considered negative. This is because work and positive energy must be provided to separate two particles attracted to one another in such a way and given that it must be assumed that the attractive field responsible for this interaction would contain no energy at all when the particles are separated by a distance that tends to infinity (in the context where the strength of the field associated with a long range interaction decreases in proportion with the square of the distance, so that it must in effect be null when this distance is infinite) then we must conclude that the energy initially contained in the same attractive force field when the particles were near one another was actually negative (so that adding positive energy can produce a null final value). This conclusion is undeniable given that it is actually observed that the energy of a bound system formed of many interacting particles is lower than the sum of the energies of those particles when they are free.

Thus, the energy contained in an attractive force field must definitely be considered negative, as this energy is required to provide the negative contribution that reduces the energy of the whole bound system. The additional energy that was present before the formation of a bound system is in fact released (through the emission of radiation for example) when the system is created, but except for the additional negative energy contained in the attractive force field the system is identical, in terms of its matter particle content, to what it was initially and therefore we definitely need the negative energy. This is made more obvious when we consider larger systems like those bound by the gravitational interaction. It was shown in effect that even a system as large as the Earth-Moon system has an asymptoti-

cally defined total mass (providing a measure of its total energy) which is smaller than that of its constituent planets (when it is possible to neglect any contribution which would normally be attributed to the presence of dark matter) and observations confirm this prediction. Therefore, it is clear that the energy contained in the gravitational field maintaining the two planets together must be negative.

What is crucial to understand regarding the situation described above, however, is that even if we must acknowledge the existence of a well-defined negative contribution to the energy of some physical systems that diminishes their total energy, it is again not possible to measure that energy directly and it can merely be deduced to occur from the behavior of the positive energy subsystems which are submitted to the attractive interaction. Here also, the negative energy must be associated with virtual particles, namely the interaction bosons that mediate the interaction, and cannot be measured independently from the total energy of the bound systems which usually remains positive. It is simply not possible to isolate the attractive field of a bound system from its positive energy sources and this is true for systems of any size. It would nevertheless certainly be a concern if the negative biding energy of a system made of positive energy components could become so negative as to make the total energy of the bound system itself negative. Once again, however, it was shown that there are unavoidable theoretical constraints on the values that observable total energy can take. It was shown [9, 10, 11, 12, 13, 14, 15, 16, 17], concerning the gravitational interaction in particular, that the energy of matter (everything except gravitation) plus that of gravitation is always positive when the dominant energy condition is assumed to be valid, which actually amounts to assume that the energy of the component particles is itself positive. If we compress positive energy matter too tightly it simply collapses into a black hole of minimum surface area and maximum energy density before the magnitude of the growing negative gravitational potential energy becomes larger than the positive energy of the matter. Thus, positive energy matter cannot turn into negative energy matter through an increase of negative gravitational potential energy.

What must be retained from the previous considerations, therefore, is that even though it is often present, negative energy seems to never be measurable. But this conclusion is valid merely under the condition that we are dealing with situations where matter was already in a positive energy configuration to begin with. It must be clear, however, that we still have no argument to rule out the possibility that there may exist configurations where the component

particles themselves would have negative energies and for which there would exist constraints similar to those unveiled here enforcing the *negativity* of energy.

In a previous section of this chapter I mentioned that it is desirable from a certain viewpoint to consider antiparticles as propagating negative energy backward in time. Indeed, if antiparticles are propagating backward in time, as the reversal of their non-gravitational charges clearly suggests, then they *must* have negative energy relative to the direction of time in which they are propagating (which is the past), so that relative to the opposite direction of time (which is the future) they would still appear to have positive energy, as required. In fact, it was discovered a long time ago by Paul Dirac (when he achieved his unification of special relativity and quantum theory) that there is a mathematical requirement for the existence of negative energy states. Indeed, it turned out that in order to obtain Lorentz invariant equations for the wave function one had to sacrifice the positivity of energy. After having considered various possible interpretations for what in nature could possibly correspond to those negative energy states Dirac concluded that it required the existence of a new category of particles, the antimatter particles, which would consist of holes in a filled distribution of such negative energy matter. But despite the fact that it was later found that antiparticles do exist, as he predicted, Dirac's solution to the problem of negative energy states was never considered fully satisfactory.

Antiparticles were eventually described by Feynman as particles propagating backward in time, which allowed to fulfill the mathematical requirements imposed by the existence of the negative energy states (by providing an interpretation for those transitions which were predicted to involve a reversal of energy) without requiring the presence of the filled negative energy continuum. But in the process it seems that the discovery that particles could actually occupy negative energy states, which appeared to be implied by the original developments, was somehow forgotten and lost in the details of the proposed solution. This indifference was probably justified by the fact that antiparticles could still be considered to have positive energy for all practical purpose. But what is usually unrecognized is that while attributing a positive energy to antiparticles may appear more 'reasonable' than assuming that those particles propagate negative energy backward in time, such a choice would actually imply that it is the particles themselves (by opposition to antiparticles) which must then be considered to carry negative energy

backward in time, because it must be either that or the opposite. This is what the subtleties of the quantum mechanical definition of energy seems to require that was not apparent classically.

The reluctance to recognize the true physical significance of negative energy states is probably also in part a consequence of the apparently insurmountable difficulties which would be associated with the possibility for particles to occupy those physically allowed states. First of all, it is certainly not desirable from a theoretical viewpoint to assume that antiparticles would be submitted to anomalous gravitational interaction as a consequence of propagating negative energy backward in time, because it was demonstrated some time ago [18] that if, for any reason, antimatter was to be found experiencing repulsive gravitational interactions with ordinary matter we would run into a number of problems ranging from violations of the conservation of energy and up to the undesirable and unlikely (from a theoretical perspective) possibility of producing perpetual motion machines. But an analysis of the arguments presented against the possibility for anomalously gravitating antimatter has led me to conclude (for reasons which will be explained later) that the problem really has to do merely with the possibility for antimatter 'as we know it' to experience what we may call antigravity. It cannot be considered to mean that matter in a true negative energy state (propagating negative energy relative to future directed time intervals) could not exist and experience anomalous gravitational interactions with ordinary matter without violating the principle of conservation of energy or the second law of thermodynamics, because matter in such a negative energy state may also by necessity have different properties from those already known to characterize antimatter, in particular with regards to non-gravitational interactions.

Nevertheless, most people today seem to consider that the developments that followed the introduction of the early theory of relativistic quantum mechanics and which gave rise to modern quantum field theory have eliminated the problem of negative energy states, which can now be considered a mere artifact of the former single particle theory. Thus, the predicted negative energy states would simply be unphysical solutions that must be discarded as irrelevant to physical reality. But it must be clear that this is indeed what we are doing here. We are rejecting the possibility that a particle could be found in a whole set of states that are allowed by the most basic equations without providing any justification as to why those states should be forbidden. Indeed, upon closer examination it becomes clear that if 'true' negative energy states do not explicitly arise in quantum field theory it is not because

the structure of the theory forbids them, but simply because we *choose* to ignore those solutions to start with and then integrate that choice into the formalism. More specifically it turns out that what prevents negative action particles from showing up in quantum field theory is merely a choice of boundary conditions for the path integrals that provide the probability amplitude for transitions involving particle trajectories in spacetime. There are several possible choices for expanding those integrals which all constitute valid solutions of the equations of the theory, but only those solutions propagating positive frequencies forward in time and negative frequencies backward in time are usually considered to be physically significant, while the solutions propagating negative frequencies forward in time and positive frequencies backward in time, which are also valid from a mathematical viewpoint, are systematically rejected. But this actually amounts to retain only the positive action portion of the theory, while ignoring all transitions involving negative action particles. There is no other origin for the often mentioned conclusion that quantum field theory does not involve negative energy matter. It is by our very choice that we reject all transitions involving negative action particles.

In order to make the choice of boundary conditions responsible for the absence of negative action particles in quantum field theory more acceptable it is sometimes suggested that the negative energies predicted by the single particle relativistic equations are simply transition energies, or differences between two positive energy states and there is obviously no reason why those variations could not be negative if they can be positive. But no explanation has ever been provided for why the same reasoning could not be applied to the energy states themselves, which are also energy differences given that the energy of a particle is always defined in relation to the zero level of energy associated with the vacuum in which it propagates. There is no justification for this arbitrary distinction between transition energies and particle energies, except for the satisfaction that is obtained by the physicist in having easily disposed of an embarrassing problem. It may of course be argued that there is nothing wrong with those methods, given that they appear to be validated by experimental results. Indeed, we have never observed interferences by negative action particles into the outcome of any experiment conducted at any level of energy and to any degree of precision. But I would like to emphasize that this still doesn't constitute an explanation for the absence of negative action particles.

Thus, the problem I have with the modern approach to quantum field

theory is that the formalism is generally introduced in a way that encourages us to believe that after all no particle is actually propagating backward in time with negative energy and that a positron is really just another particle, identical to the electron, but with an opposite electrical charge. However, this viewpoint does not only complicate things unnecessarily as a consequence of rejecting the possibility for electrons and positrons and all other particles and their related antiparticles to actually consist in the same particles observed from different perspectives, it is also completely ignorant of the requirement of a relational definition of any physical attribute dependent on the fundamental time-direction degree of freedom. But if we choose to recognize the validity and the greater value of the viewpoint defended here and according to which antiparticles are really just ordinary particles propagating backward in time, then we must accept that there definitely exist in nature particles which are known as carrying negative energies and if the arguments provided above concerning the arbitrariness of the current restrictions imposed on the propagation of those negative energy states are valid, then we would have to conclude that there should necessarily also exist particles with such energies propagated forward in time and which could be submitted to anomalous gravitational interactions in the presence of ordinary matter.

1.4   The negative mass concept

When discussing the issue of negative mass what must first of all be understood is that if the physical property of mass is to have any polarity associated with it, such that we could attribute to mass either a positive or a negative sign, then this polarity must be directly related to the sign of action, that is, to the sign of energy relative to the positive direction of time. This is because, as I previously emphasized, the sign of action is the only physical property from which the attractive or repulsive character of the gravitational interaction between two bodies could depend. We may thus attribute positive mass to a positive action particle and negative mass to a negative action particle. Mass being a Newtonian concept its polarity must be determined in relation to a particular Newtonian gravitational field. From this viewpoint the sign of mass of a given particle could in effect be understood as determining the response to the gravitational field of a given source, in the sense that it would determine the *direction* of the gravitational force exerted on such a particle. If we may consider the gravitational field of the source (represented

by a vector in Newtonian mechanics) to be uniform, then only its own direction or polarity (which we may assume to be dependent merely on the sign of mass of the source when its position is assumed to be fixed) would be decisive in determining the kind of response experienced by a given type of mass submitted to it. Equipped with such a definition we can meaningfully discuss the problem of the gravitational interaction of negative action particles with positive action particles and with themselves as the problem of the gravitational interaction of positive and negative masses. This will allow us to better grasp the significance of the assumptions that will form the basis of the new interpretation of negative energy matter which I shall propose and therefore, also, to gain better confidence in their validity, even in the more appropriate context of a general relativistic theory.

If we may agree on those requirements, then I think that what must emerge is that if it is indeed important to have a well-defined concept of negative mass then it also seems that such a negative mass must be negative in all respects. That there could be a difference between the sign of gravitational mass and the sign of inertial mass is usually considered to be forbidden merely by the general theory of relativity which is in effect founded on the principle of equivalence which requires the equality of gravitational and inertial masses. However, I think that if this hypothesis is justified it is not because our concept of mass polarity must comply with some perceived requirement from general relativity theory, but because it would not be acceptable to attribute mutually exclusive values to a single unique physical property. Thus, I do believe that the mass of any particle or body should be either definitely positive or definitely negative (but still in a relational way), regardless of whether we are considering gravitational mass or inertial mass, if the concept itself is to have any consistent physical meaning. But unlike most theorists I do not consider that this requirement must be assumed to imply the kind of behavior that is usually attributed to negative mass matter, where gravitational repulsion is an intrinsic property of this type of matter itself, independently from the sign of mass of the matter with which it is interacting. This is indeed the conclusion I was able to draw based on the outcome of the previously discussed analysis of the constraints imposed by a relational definition of the sign of energy, for reasons I will now explain.

The difficulty I originally met when I first began to explore the possibility that inertial mass could be reversed along with gravitational mass when we are dealing with negative mass matter is that if both the gravitational mass and the inertial mass are to be negative at once then it seems that there could

occur situations where the principle of inertia would be violated (I will explain what motivates this belief below). I was able to understand, however, that this is merely a consequence of the inappropriateness of current assumptions regarding what we should expect to be the behavior of matter with both a negative gravitational mass and a negative inertial mass. Actually, despite the fact that it is usually taken for granted that we know for sure at least what the behavior of matter with positive mass is, because we routinely observe gravitational phenomena involving this kind of matter and there can be no mistake here, I will explain that this is not entirely the case and that there is still much confusion as to even what we should expect concerning the response of positive mass matter to a concentration of negative mass. Currently it is assumed that given that positive mass matter gravitationally attracts all matter and resist the action of any force exerted on it, then this must be an intrinsic property of such positive masses. On the other hand, it is usually assumed that two choices exist for what could possibly characterize the behavior of matter with a negative mass. The situation we have right now is thus the following.

First of all, we must assume that *gravitational* mass is indeed negative when mass is reversed. This allows to obtain gravitational repulsion when only the mass of the source (the active gravitational mass) is negative, because it reverses the polarity of the Newtonian gravitational field to which any passive gravitational mass is submitted and therefore should at least reverse the force exerted on positive mass bodies. But once this is recognized it is usually considered that two possibilities actually exist for a negative mass particle submitted to a given gravitational field, depending on whether *inertial* mass is assumed to remain positive or is itself also negative. Here the inertial mass of a particle is assumed to determine the response of that particle (actually the direction of its acceleration) to any force, including a gravitational force, while the gravitational mass of the same particle is assumed to determine both the polarity of the gravitational field it produces and the response of the particle to a gravitational force. If we were to agree with those assumptions then we would have to conclude that a negative gravitational mass particle with a negative inertial mass, should actually respond normally to any gravitational force field (because the nature of its response is changed twice, once by the reversal of its inertial mass and once by that of its gravitational mass) while its response to non-gravitational forces would be reversed (same force, opposite acceleration), as current assumptions concerning the effects of a reversal of inertial mass would imply. But we must

also keep in mind that the fact that this kind of matter would respond normally to gravitational force fields would, under the current assumptions, still mean that it is repelled by matter of the same type, because the gravitational field produced by such matter is also assumed to be reversed. Thus, such negative masses would repel masses of all signs, be repelled by other negative masses and be attracted to positive masses, still under the hypothesis that the above stated commonly accepted assumptions are valid. Given that it is usually considered that in a general relativistic context all mass (gravitational and inertial) must be negative, this is the choice that is usually retained as defining the behavior of negative mass matter if it could exist.

But despite the support that is usually granted to such a conception of negative mass or negative energy matter I think that enormous problems would arise if it was retained as a valid proposal. Some of those problems, involving black holes and the second law of thermodynamics, will be discussed later, but even if we remain at the level of classical Newtonian dynamics we can readily identify one very serious problem which is that the existence of such matter would allow violations of the principle of inertia (considered as a generalization of Newton's first law) or the very idea that no physical system can accelerate without work being done on it by an external force. This is because indeed, as stated above, from the current viewpoint a negative mass body would both repel positive mass bodies and be attracted to them. Such a combination of features could then give rise to unlikely phenomena like pairs of opposite mass bodies chasing one another and in the process accelerating to infinite velocities, still without any external energy input [19]. The fact that energy would in principle be conserved under such conditions (because the energy gained by one of the bodies would be opposite that of the other) is no consolation, because we are dealing here with a much more serious and basic violation.

Indeed, the problem I see is that there would be no equal and opposite force to that applied on a given body that could be attributable to its assumed interaction with the other body and this would be a violation of the principle of action and reaction (Newton's third law), which is one requirement that in all fairness we should recognize as being even more fundamental than that of conservation of energy, because if it does not rigorously apply then absolutely anything could occur and under such conditions we could not give much of even the principle of conservation of energy. However, I think that what those observations show is not the unphysical nature of negative mass, but merely the ineffectiveness of the traditional approach to describe the

behavior of this kind of matter. It is important to mention, by the way, that even though this hypothetical situation of accelerating opposite mass pairs has been described by other authors in the past, none of them has ever recognized that what it actually demonstrates is the inconsistency of the current notion of negative mass, which I believe is illustrative of the state of denial in which most people remain concerning the possibility that there could actually exist negative mass matter.

What is also significant concerning the unlikely phenomenon described above is that it would necessarily be the positive mass bodies that would be chased in this way, while the negative mass bodies would inevitably be those trailing them. But isn't it strange indeed that there should be such a clear and decisive distinction between what constitute the role of positive masses and what constitute that of negative masses? Doesn't it seem like there is something wrong with such a hypothetical phenomenon? Shouldn't we only be allowed to define the property of gravitational attraction and repulsion in such a way that we could not observe such mass-sign-distinguishing be- havior? What I have understood is that the unease we may experience in face of the strangeness of such phenomena is in fact justified. Indeed, it does not just seem like there is something wrong here, because what we have just described is actually the perfect example of an attempt to distinguish a physical property (the positivity of mass or the attractiveness of gravitation) despite the absence of any reference in the physical universe to which that arbitrary distinction could be related, which violates the very basic require- ment of relational determination of physical attributes discussed above. The mistake which is made by assuming the validity of the traditional viewpoint is that we suppose that we can define attraction and repulsion in an absolute (non-relative) manner such that one kind of mass always attracts all kinds of masses regardless of their polarity and another always repels all masses, still regardless of their sign, as if attractiveness and repulsiveness were intrinsic aspects of one and the other type of mass.

However, if the sign of mass is to be considered a meaningful physical property of elementary particles then it must be taken to indicate that there can be a reversed or opposite value to a given mass and this reversed value can be considered to be reversed merely in relation to a non-reversed mass and to nothing else. A mass cannot be considered to be reversed with respect to an absolute point of reference lacking any counterpart in the physical universe. Therefore, if a gravitational field is to be assumed repulsive as a consequence of the reversed (negative) sign of the mass of the matter that is

the source of the field then this gravitational field should be repulsive only for an unchanged (positive) mass particle and not with respect to other negative masses. It would be incorrect to assume that the attractive or repulsive nature of gravitational fields depends solely on the sign of mass of the source itself, because no distinction exists for the sign of a mass other than its sameness or oppositeness compared to that of another mass. That does not mean that the field itself must be assumed to change as a consequence of the reversal of the sign of mass of the particle experiencing it (even though that may be one way to describe things if other conventions are adopted for the sign of mass itself as we will see later), but merely that the response of a negative mass particle to a given gravitational field must be reversed in comparison to the response we would expect from a positive mass particle submitted to the same field, despite the associated reversal of the inertial mass of such a particle. If that was not the case, then I think that we would have to conclude that negative mass is, in effect, forbidden.

If the incorrect hypothesis on which the traditional approach is based regarding the effect of a reversal of inertial mass nevertheless allows to successfully (from my viewpoint) predict that a positive mass would be repelled in the gravitational field of a negative mass, it is simply because we assume the right inertial properties for the positive mass matter submitted to the gravitational force of the negative mass. Thus, the positive mass responds in the appropriate way to the gravitational force exerted by the negative mass which is correctly assumed to be a repulsive force given that the gravitational field produced by the negative mass is necessarily opposite that which would be produced by a positive mass of similar magnitude located in the same position. The problem is that given that it seems that we cannot expect the same kind of behavior from a negative mass submitted to the gravitational field of a positive mass, then it would appear that the behavior of both positive and negative masses is the consequence of some predetermined property of absolute attractiveness and repulsiveness (that cannot be related to any property of the source defined with respect to a property of the matter with which it interacts) associated with the gravitational fields emanating from positive and negative masses respectively.

The difficulty to which the traditional interpretation gives rise is also made apparent when we consider the case of a negative mass in the gravitational field of another negative mass, given that now the negative mass would be repelled by the same negative mass matter (because the gravitational force is unchanged but the response to this force would be reversed),

while on the basis of the relational definition of mass sign there should be no difference between this case and that of a positive mass in the gravitational field of another positive mass (which is symmetric to the other case under exchange of mass signs). The appropriate outcome could only be obtained if in addition to the assumption regarding the nature of the gravitational force between two negative mass bodies it is also assumed that the reversal of the inertial mass of the negative mass body submitted to this force actually changes nothing to the response of that body to the attractive force of the other negative mass body. Thus, the problem of the absoluteness of the attractive or repulsive nature of the gravitational field arises as a direct consequence of current assumptions regarding the effect of a reversal of inertial mass. It is only in this context that the direction of the Newtonian gravitational field associated with a concentration of matter of positive or negative mass sign acquires an absolute meaning and is not merely dependent on a convention as to what should be the sign of mass of the matter that is the source of this field.

Even if merely as a consequence of the previously discussed considerations regarding the relative nature of the sign of energy (as dependent on the direction of propagation in time of a particle) and the purely conventional (subject to an arbitrary coordinative definition) significance of the sign of action it would appear that a consistent notion of negative mass would require that it is the relative difference or absence of difference between the mass signs of two gravitationally interacting bodies that determines the attractive or repulsive character of this interaction, so that two negative mass bodies should be submitted to the same mutual gravitational attraction experienced by two positive mass bodies and would also repel ordinary positive mass bodies and be repelled by them, unlike is usually assumed. But the fact that it is often not even fully understood that negative mass should in effect be associated with negative action is illustrative of the confusion that surrounds the whole question of negative energy and gravitational repulsion, because there should be no doubt that if it is possible for the sign of mass of a given body to be negative in some way, then this would necessarily have to occur as a consequence of the fact that this body has negative energy, or more precisely negative action. In any case, if the traditional viewpoint allows predictions that violate the expectations of a relational definition of mass sign it is precisely because it allows to assume that there can be an absolute character of attractiveness or repulsiveness associated with a given sign of

mass. To be fair, I must acknowledge that some authors did suggest in the past that the gravitational interaction should perhaps be repulsive between bodies of opposite mass signs while it would be attractive between negative mass bodies (just as it is between positive mass bodies), but simply on the basis of the fact that the sign of the gravitational force that is obtained by reversing the sign of one of the masses in Newton's equation for universal gravitation would itself be reversed, while it would be unchanged if the signs of the two masses were together reversed.

But even though it is not necessarily wrong to suggest that the repulsive or attractive nature of the gravitational interaction is determined by the sign of the force in Newton's equation for universal gravitation, it is only when we realize that the sign of mass must be related to the sign of action that we can begin to understand why it is that there should be a symmetry under exchange of positive and negative masses. This is because, as I previously mentioned, positive action states are related to negative action states by a simple convention regarding the sign of energy and that of time intervals, so that the sign of action is itself a purely relative notion. There must consequently be a symmetry under exchange of positive and negative action matter, which would then require the behavior of positive masses in relation to themselves and in relation to negative masses to be similar to that of negative masses in relation to themselves and in relation to positive masses. I may add that in such a context it appears that the suggestion that if negative mass bodies have never been observed it is perhaps simply because they do not assemble themselves into larger masses (as a consequence of their assumed absolute gravitationally repulsive nature) cannot be valid and if negative mass matter exists, then alternative arguments would have to be proposed to explain this absence of observational evidence. Later in this chapter I will indeed explain how it is possible to reconcile the apparent absence of large scale concentrations of gravitationally repulsive matter with a more consistent notion of negative mass matter.

The contradictions of the traditional conception of negative mass or negative energy matter can be illustrated by using a rarely discussed thought experiment. It has in effect been proposed that the sign of energy of a negative mass particle could be determined by measuring the energy lost or gained while raising or lowering the particle in the gravitational field of some large object. Now, according to the traditional conception if we were to raise a negative mass body in the gravitational field of a positive mass object like a planet we would have to produce work and exert a force directed down-

ward because the inertial mass of the body is negative, which according to the traditional viewpoint means that it responds perversely to the applied force. But then it is also the case according to this same viewpoint that the gravitational force exerted by the planet on the body should be attractive, because the planet has positive mass. Thus, we would be in the situation where we would have to exert a force downward to raise a negative mass body in the gravitational field of a planet that exerts an *attractive* force on that body. I do not know to what extent people actually believe in the validity of such a conclusion, but I think that faced with such absurdities one has to come to realize that the contradictions involved are a clear indication that the traditional assumptions regarding the behavior of negative mass or negative action matter are incorrect and that a better interpretation of what such a state of matter may involve is required.

Despite the fact that the question of the validity of the traditional conception of negative mass matter had never been clearly analyzed before, it is no doubt the general feeling that there is something wrong with the possibility of observing phenomena of the type described above (including that where pairs of opposite mass bodies accelerate without any external force being applied on them) which is responsible for having transformed the idea of negative energy or negative mass matter into the synonym of nonsense it has become in the minds of so many researchers. But, is negative mass really to blame here or could it be that we are not considering the right possibility? There is of course, even under the conventional assumptions regarding the response of negative mass particles to applied forces, another possibility which is that when gravitational mass is negative, inertial mass may remain positive for some reason. Of course that would not only appear to contradict the equivalence principle, as is already understood, it would also, if I'm right, itself be nonsense, as we would have to assume that one single physical quantity related to one single particle (the mass of that particle) is at once both positive and negative for the same observer. The latter problem has never been discussed, but I think that it is actually the strongest argument one can make against this second possibility. We may nevertheless begin by exploring the consequences of such a choice.

Under the same commonly held assumption to the effect that the response of a particle to any force is dependent on the sign of its inertial mass we would have to conclude that a negative gravitational mass body to which a positive inertial mass would be attributed would respond anomalously (in

comparison to the response expected of a positive mass) to any gravitational force field (because the nature of the response is changed only once by the reversal of its gravitational mass), while its response to non-gravitational forces would be unchanged (same force, same acceleration), because the inertial mass remains positive or unchanged in comparison with that of positive mass bodies. Therefore, if material bodies were to exist that would be made of such negative mass matter they should, from the traditional viewpoint, gravitationally attract one another (as do positive masses), repel positive mass bodies and also be repelled by those same positive mass bodies. As a consequence, we would observe no violation of the principle of inertia in this case and also no acceleration without work. If this behavior was to be observed it would in fact be possible to exchange all positive mass bodies with negative mass bodies and vice versa and no apparent change in the phenomenology of the gravitational interaction would be detectable, because gravitational repulsion would only occur when there is a difference in the signs of the *gravitational* masses which are interacting. Thus, from a purely phenomenological viewpoint there would be equivalence between positive and negative mass bodies.

Given the previous discussion regarding the necessity of a relational determination of the sign of energy, which would here be a requirement for the relational determination of the sign of mass, this situation would appear more appropriate, because indeed it would be impossible in principle to differentiate any intrinsic property of gravitational attraction or repulsion and only the difference or the equality of the signs of *gravitational* mass of two particles would be physically significant. The problem that most people would have with this possibility, however, is that it would explicitly violate the equivalence principle, because positive and negative gravitational masses would respond differently to a given gravitational field, produced by a given matter distribution, even if they are located in the same local inertial frame of reference.

But I think that even before we consider the issue of the apparent incompatibility with the principle of equivalence we must first of all ask how could it be determined which of the two types of matter would indeed have the *inertial* mass opposite its gravitational mass? And then it is obvious that this question could never be settled (because we could never decide which type of matter actually has a negative gravitational mass) and yet in such a context this would be a highly pertinent question, as we do assume a physical difference analogous in this respect to an absolute distinction between

positive and negative mass bodies. Indeed, why would the inertial mass remain positive when the gravitational mass is reversed. It is only confusion to pretend that there are multiple aspects of mass and that each of those independent mass properties can have a different sign. An electric charge is either positive or negative and mass clearly defined as the charge associated with the gravitational interaction must also be either positive or negative and this is actually all that the equivalence principle requires I believe. In this context I think that we would be right to object trying to save the principle of inertia by assuming that negative masses could at once also have a positive inertial mass, because this would indeed violate the equivalence principle, not because different masses could accelerate in different directions in a gravitational field, but to the contrary because indeed the *same* inertial masses could actually respond differently to a given gravitational field, which would then really mean that there definitely cannot be equivalence between a Newtonian gravitational field and acceleration and this would indeed be a problem for relativity. Clearly there is still something wrong, even with the second alternative that is traditionally considered for the attribution of negative mass.

The preceding discussion should then have made clear the fact that there are two issues regarding negative mass. First, if we accept the requirement for a relational definition of the attractive and repulsive character of a gravitational field, then we must conclude that the currently favored assumption for what would be the behavior of negative mass bodies, having at once negative gravitational mass and negative inertial mass, is incorrect, because, as I explained, it would involve absolutely defined properties of attractiveness and repulsiveness that would not depend merely on the difference or equality of the signs of the interacting masses. But if we consider the other traditionally considered (but not favored) possibility for the definition of negative *gravitational* mass, we may obtain the required relational definition of gravitational attraction and repulsion, but as I have explained a distinct problem would arise.

Indeed, under such conditions the appropriate behavior expected of negative mass matter would have to be that which we currently assume to be possessed by particles with a contradictory definition of their mass sign, which is not only objectionable on the basis of consistency, but which also violates the equivalence principle in a way that cannot possibly be allowed (same mass, different response) if relativity theory is to be retained as a valid theory (if we were to accept this possibility then there would be no reason

why relativity itself should still be required). Arguing that the problem here is with the notion that mass is at least in part the same, while this identity of mass signs actually applies merely to a different property of mass which we would call inertial mass, so that the 'real mass', which we would call the gravitational mass, could be different, would in my opinion not just be confused, it would be nonsense. What is positive cannot also at the same time be negative if this polarity is to have any meaningful physical significance. Mass is not an abstruse, complicated property with multiple independent and yet interrelated aspects, it is the gravitational charge and even though the stress-energy tensor replaces mass as the source of gravitational fields in a general relativistic context, the lessons learned here are still valid and significant even in the context of the modern theory of gravitation.

It took me some time to realize that the problems we are dealing with here (if we are willing to recognize that the whole question of identifying the properties of negative energy matter is not itself insignificant) originate from what is usually assumed concerning the response to any force field in the case of a body with negative inertial mass. It is only after a rather long process of getting to understand the meaning of the phenomenon of inertia that I was finally able to gain the insight required to solve the problem of identifying the actual properties of negative mass matter in the context where we consider it a consistency requirement to impose on such matter that it should have both a negative gravitational mass and a negative inertial mass. Keep in mind that this explanation will be easier to grasp when the consequences of the integration of such a concept of negative energy matter to the modern theory of gravitation will have been more thoroughly explored. Basically what must be understood is that the direction of the equivalent gravitational field experienced by a given mass in a frame of reference in which it is accelerating, even in the absence of nearby matter inhomogeneities, is in fact dependent on the sign of the mass that is accelerating. As a consequence of this hypothesis the inertial *force* associated with a given acceleration is left invariant even if the sign of inertial mass is itself reversed along with the gravitational mass for a negative energy particle.

In order to appreciate the following discussion at its true value it is essential to remember that relativity theory does imply in effect that there exists a Newtonian gravitational field exerting a gravitational force on a positive mass body which is accelerating relative to a local inertial frame of reference, even far from any large mass. The existence of the inertial force associated with this equivalent gravitational field is what allows a dynamic (by oppo-

sition to static) equilibrium to occur when an external force is applied on a body which gives rise to an acceleration. Indeed, in the accelerated frame of reference relative to which a positive mass body submitted to an external force does not accelerate a gravitational force is present which balances the applied external force and this is what explains that there is no acceleration of the body relative to this particular (accelerated) frame of reference. In fact, the equivalent gravitational field is a general feature of acceleration and is present in any accelerated frame of reference, but in the absence of an external force to balance the associated inertial force the equivalent gravitational field only serves to determine the local inertial frame of reference associated with free fall motion.

Indeed, given that the force associated with the equivalent gravitational field is a gravitational force we must conclude that when the force responsible for the acceleration is itself gravitational we are actually in a situation where there would appear to be no force at all. It is therefore possible to assume that what determines the local inertial frames of reference relative to which a positive mass experiences no gravitational force is the local matter distribution which is the source of the applied gravitational forces which are balanced by the inertial force which would otherwise be present relative to those reference systems (this is the essence of the insight that led to relativity theory). In any case it is clear that the inertial force attributable to an equivalent gravitational field is always directed opposite the direction of the external force which gives rise to the corresponding acceleration for a positive mass body and this means that the direction of the equivalent gravitational field experienced by a positive mass body is opposite the direction of its acceleration, that is, opposite the direction of acceleration of the frame of reference relative to which this equivalent gravitational field exists. But what would occur if we had a negative mass body in place of a positive mass body?

First of all, it must be clear that the gravitational force $\boldsymbol{F}_g = m\boldsymbol{g}$ on a particle of mass m attributable to a given matter distribution would be reversed if the mass of the particle was reversed, because the Newtonian gravitational field vector $\boldsymbol{g}$ at the particle's position would be left unchanged (because the matter distribution that is the source of the field does not change), while the sign of mass of the particle experiencing the field would be reversed. Now the problem usually is that when we want to determine the response of a particle to some gravitational force $\boldsymbol{F}$ using Newton's second law $\boldsymbol{F} = m\boldsymbol{a}$, if the mass of the particle is reversed (negative) then the

resulting acceleration $\boldsymbol{a}$ would appear to have to be opposite that experienced by a positive mass submitted to the same force (the acceleration would be in the direction opposite that of the applied force). This is the traditional conception regarding negative mass. But if we consider things in a more general context, where Newton's second law would be an equation expressing the equilibrium between external forces $\boldsymbol{F}_{ext}$ and the inertial force $\boldsymbol{F}_i = m\boldsymbol{g}_{eq}$ effected by the equivalent gravitational field $\boldsymbol{g}_{eq}$ associated with a given acceleration, then we may write $\boldsymbol{F}_{ext} + \boldsymbol{F}_i = 0$ or $\boldsymbol{F}_{ext} = -\boldsymbol{F}_i$ so that for example if the external force is gravitational $\boldsymbol{F}_{ext} = \boldsymbol{F}_g = m\boldsymbol{g}$ then we would have $m\boldsymbol{g} = -m\boldsymbol{g}_{eq}$ and this means that the equivalent gravitational field $\boldsymbol{g}_{eq}$ is usually opposite both the applied gravitational field and the acceleration, because in the present case we also have $\boldsymbol{F}_{ext} = m\boldsymbol{a}$, which means that $m\boldsymbol{g}_{eq} = -m\boldsymbol{a}$ for the considered positive mass m at least.

But would the equivalent gravitational field experienced by a negative mass particle really be directed opposite the direction of its acceleration as is the case for a positive mass particle? To that question I think that, contrarily to what is usually assumed implicitly, we would have to answer that this cannot be the case. I will explain that in fact the equivalent gravitational field $\boldsymbol{g}_{eq}^-$ that would be experienced by a negative mass particle accelerating in a given direction away from any local matter inhomogeneity is the opposite of the equivalent gravitational field $\boldsymbol{g}_{eq}^+$ that would be experienced by a similar positive mass particle with the same acceleration under the same conditions, so that we have $\boldsymbol{g}_{eq}^- = -\boldsymbol{g}_{eq}^+ = -(-\boldsymbol{a}) = \boldsymbol{a}$ for a negative mass particle and given that we still have $\boldsymbol{F}_{ext} = -\boldsymbol{F}_i = -m\boldsymbol{g}_{eq}^-$ it means that $\boldsymbol{F}_{ext} = -m\boldsymbol{a}$ when the mass m is negative. If this is correct then it would mean that the acceleration which a negative mass particle would experience as a result of the action of a given force would actually be the same as that which would be experienced by a positive mass particle submitted to the same force (not the same force field but really the same force), even if the mass, including the inertial mass, is indeed negative. The validity of this conclusion depends on only two assumptions. First, the proposed generalized Newton's second law (explicitly involving inertial forces instead of accelerations) must be considered more fundamental than the original formulation involving accelerations, so that the equilibrium it describes is really between forces and not merely between a force and an acceleration. Secondly, it must be assumed that the equivalent gravitational field associated with a given acceleration is reversed when the mass is reversed.

If the preceding conclusions are accurate it would appear that the fact

that Newton's second law was always observed to work in its original form, that is, when the equivalent gravitational field is implicitly considered to be opposite the acceleration, is merely a consequence of the fact that it has only ever been verified to apply using positive mass matter. But what is it indeed that might allow one to assume that the equivalent gravitational field would be reversed (would be directed in the same sense as the acceleration) for an accelerating negative mass particle in comparison to what it would be for a similarly accelerating positive mass particle? To understand what is going on we may consider the example of Einstein's elevator experiment. Indeed, we are allowed by the equivalence principle to assume that the effects observed inside an elevator accelerated in the vacuum away from any local matter inhomogeneity could also be explained by assuming that the elevator is not accelerating in the same vacuum (relative to the local inertial frame of reference which would exist in the absence of any local matter inhomogeneity), but that a large mass, not originally present in this vacuum, is now located beneath the elevator (in the direction opposite that of the originally assumed acceleration). Thus, it seems that acceleration relative to a local inertial frame of reference always gives rise to an equivalent gravitational field similar to that which we would normally attribute to the presence of a local concentration of matter. We may then define an *equivalent source* to be the matter distribution which would give rise to the equivalent gravitational field experienced by an accelerated body if the presence of this field was not merely the consequence of acceleration.

Now, if we are allowed to assume that the equivalent gravitational *field* associated with the inertial gravitational force is actually reversed when the mass of the accelerated body is itself reversed (even without speculating about what the phenomenon of inertia might actually involve), it is simply because we can expect that the sign of mass of the equivalent source associated with the equivalent gravitational field experienced by a negative mass body should itself be reversed. There should be no question in effect that if an accelerating positive mass observer is allowed to assume that the equivalent gravitational field she experiences is actually attributable to the presence of an equivalent source with *positive* mass located in the direction opposite her acceleration, then a similarly accelerating negative mass observer should himself be allowed to attribute the equivalent gravitational field that he would experience to the presence of some equivalent source with *negative* mass also located in the direction opposite his acceleration, otherwise we would have a way to determine in an absolute fashion, the positivity of mass.

Indeed, if it was always an equivalent source with positive mass (located in an invariant position relative to the accelerating body) that gave rise to the equivalent gravitational field, we could simply accelerate an observer of any mass sign and measure the equivalent gravitational field experienced by this observer, which could then be identified as the gravitational field attributable to a positive mass in the assumed position. Therefore, any gravitational field exerting on a given body a force such as that which was observed could be identified as the gravitational field of a positive mass independently from the mere difference or equality between the polarity of the mass producing the field and that of the particle experiencing it. But this is a violation of the above discussed requirement of relational definition of the sign of mass. Thus, the problem with the traditional conception of negative inertial mass is that it would again allow to differentiate between positive and negative mass in an absolute (non-relative) way, this time by referring to the predefined positive mass of the equivalent source whose gravitational field should invariably be observed under otherwise arbitrary motions of acceleration.

As it turns out, an additional difficulty arises when we try to assess the response of negative mass matter to applied forces if we insist on assuming that the equivalent gravitational field associated with acceleration is an invariant property of the acceleration itself. Indeed, it is not only in the presence of an external force that the inertial force on a negative mass body would have to be in the direction of its presumed acceleration when it is assumed that the equivalent gravitational field is opposite this acceleration (as is the case for a positive mass body). The truth is that, when one recognizes the validity of the generalized form of Newton's second law, then under the inappropriate assumption that it is an equivalent source with positive mass that gives rise to the inertial force experienced by a negative mass body in an accelerated reference system, it follows that even in the absence of external forces the inertial force would have the same direction as the acceleration, which means that the negative mass body would actually accelerate in the same direction as the *accelerated* reference system itself. As a consequence, there would no longer be an equilibrium between the applied forces and the inertial force that is experienced by a negative mass body due to its acceleration, which is certainly not a desirable outcome. Thus, even if the equivalent gravitational field experienced by an accelerating negative mass body was the same as that experienced by a similarly accelerating positive mass body, this would not give rise to the kind of motion which is traditionally expected from a negative mass body.

What is important to understand in effect is that in the context of a generalized formulation of Newton's second law it must actually be imposed that there is always an equilibrium between the applied forces and the inertial force and under such conditions the acceleration to which a body with a given mass sign is submitted is determined solely by the requirement that the inertial force it experiences actually balances the applied forces. Thus, once the direction of an applied force is known the acceleration of the body submitted to this force is determined only by the condition that it does in effect give rise to an inertial force which balances the applied force. But if the equivalent gravitational field which gives rise to the inertial force is dependent on both the direction of acceleration and the sign of mass of the accelerated body then the fact that the sign of mass would be reversed would not affect the direction of the acceleration, because the equivalent gravitational field would also be reversed, which allows the inertial force associated with this acceleration to remain invariant under a reversal of mass.

Under such conditions it would not be appropriate to assume that it is the sign of mass itself which determines the direction of the acceleration, because in fact the acceleration of a body submitted to a given force is determined merely by the requirement that the inertial force experienced by such an object balances the applied force in the accelerated reference system relative to which this inertial force is present. There is no a priori justification for considering that a negative mass body with negative inertial mass should experience an acceleration opposite the applied force. This would be an incorrect interpretation of the classical equation between force and acceleration, which must be assumed to be valid only when the mass is positive. What the preceding argument shows in effect is that it would be a mistake to assume that the traditional formulation of Newton's second law also applies when the mass is negative. This equation does not apply when the mass is negative simply because the formula was not derived under the assumption that mass can be negative and was never intended to apply under such circumstances. But in the context of a generalized formulation of Newton's law and when the mass of the equivalent source responsible for the equivalent gravitational field is appropriately reversed for an accelerating negative mass body, it follows that the equivalent gravitational field experienced by such an object must itself be opposite that experienced by a positive mass body, which means that the inertial force remains unchanged, as does the body's acceleration.

If we are willing to recognize that it would be a serious inconsistency to

allow for the same equivalent source (with the same mass sign) to give rise to both the equivalent gravitational field experienced by positive mass particles and that experienced by negative mass particles, then we must also recognize that similarly accelerating positive and negative mass bodies would experience opposite equivalent gravitational fields, because those gravitational fields would arise from equivalent sources with opposite mass signs. But given that a negative mass must experience a force opposite that experienced by a positive mass of similar magnitude in response to any gravitational field, it follows that the inertial *force* would actually have the same direction for both positive and negative mass bodies accelerating in the same direction as a consequence of being submitted to the same external force (which is more constraining than requiring the same applied force *field*), even if we consider inertial mass to be reversed along with gravitational mass, as I previously argued to be necessary.

In the present context we would actually be allowed to assume that the requirement to consider that the equivalent gravitational field is reversed for a negative mass body (in comparison with the equivalent gravitational field experienced by a positive mass body with the same acceleration) is justified by the fact that it allows the dynamic equilibrium of forces on such an object to be maintained in the accelerated frame of reference relative to which this equivalent gravitational field is experienced, because if in order to meet this constraint we must consider the same inertial gravitational force to arise from the same acceleration then it means that a negative mass body would necessarily have to experience a reversed equivalent gravitational field given that its mass is indeed reversed. No circular reasoning is involved here, because those results actually follow from the mere requirement of relational definition of the sign of mass applied to the equivalent source that gives rise to the equivalent gravitational field experienced by an accelerating negative mass body.

For this argument to be valid what must be recognized is that the negativity of the inertial mass of a negative 'gravitational' mass is an independent consistency requirement, which actually amounts to assume that mass is mass and that it cannot be both negative and positive at the same time and once this is acknowledged we are allowed to also and independently conclude that just as there is not a unique sign of mass, there is not a unique equivalent gravitational field for bodies with opposite mass signs. In such a context we have no choice but to recognize that the response of a negative mass body to any applied force would be that which we ordinarily (but inappropriately)

attribute to a negative gravitational mass whose inertial mass would remain positive. Yet in the present case it would seem that the validity of the equivalence principle could be preserved to some extent, even while there are two different kinds of response to a given gravitational field, because all mass (gravitational and inertial) is now reversed for a negative mass body and only bodies with opposite mass signs must be assumed to respond differently to a given gravitational field and not bodies with the same inertial mass, which would then have constituted a real violation of the requirement of equivalence of acceleration and gravitation, as I explained before.

It is now possible to understand why it is that the inappropriate choice of a positive inertial mass in association with a negative gravitational mass would seem to agree, from a purely phenomenological viewpoint, with the independently motivated requirement of a relational definition of mass sign (given that it would allow gravitational attraction and repulsion to themselves be features dependent merely on the *difference* between the signs of gravitational mass of any two bodies). It is simply because in such a case instead of appropriately reversing the equivalent gravitational field for a negative mass accelerating in a given direction we would reverse the sign of inertial mass (which must be negative for a negative mass particle) a second time, from negative to positive again (while keeping the gravitational mass negative), which superficially would be equivalent to simply reversing the direction of the equivalent gravitational field while keeping the mass negative as required. But I must emphasize again that if that was the only possible approach to obtain consistent behavior from negative mass bodies we would in fact have to conclude that negative mass is not an appropriate concept in physical theory, because we would have to assume that a single unique physical property (what we may call the gravitational 'charge') is required to have at once and *from the exact same viewpoint* (for an observer of unchanged mass sign) two opposite values and this is clearly unacceptable.

It must nevertheless be mentioned that, as later developments will illustrate, it appears that in fact the reversal of the equivalent gravitational field is the trade-off we have to accept for keeping the value of the gravitational field attributable to a local matter inhomogeneity generally invariant while assuming that it is actually the mass experiencing it that can be reversed. But if instead we considered that the motion of a body must always be determined from the viewpoint of an observer made of matter with the same sign of energy, then it would be natural to assume that the sign of mass of the body (both inertial and gravitational) is an invariant property that may

be assumed positive definite, while it is the gravitational field attributable to a given matter inhomogeneity that is variable as a function of the difference between its energy sign and that of the observer.

From this viewpoint the equivalent gravitational field due to acceleration far from any local matter inhomogeneities would no longer be dependent on the sign of mass of the accelerating body (because the mass itself would not change), while the gravitational field due to the presence of a local matter inhomogeneity would depend on the perceived sign of energy of its sources which would become an observer dependent property, while the mass or energy of the body experiencing the fields would actually be considered positive definite. In this context there would then still be a practical (although not fundamental) distinction between an equivalent gravitational field due to acceleration far from any local mass concentration and the gravitational field due to the presence of a local matter inhomogeneity (in the absence of forces other than gravity). I will explain below what is the profound origin of this distinction and why it does not constitute an insurmountable difficulty for a consistent general relativistic theory of gravitation based on the equivalence principle.

What must be retained here is that we can still consider the direction of the gravitational field attributable to the presence of a local matter inhomogeneity to remain invariant while it is the mass experiencing it and therefore also the equivalent gravitational field experienced by this mass which may be reversed, but only at the price of changing the equations of motion which will be shown to otherwise describe the trajectories of particles submitted only to the gravitational interaction in a way that is equivalent to considering that the mass experiencing the gravitational field (due to the local matter inhomogeneity) is invariant while it is the field itself which is reversed (in comparison to what it would be if we had considered its effect on a negative mass body). Now, if we do consider the mass (both gravitational and inertial) of the particle experiencing a gravitational field to always be positive definite so that that it is the direction of the gravitational field itself which varies as a function of the *relative* difference between the observer dependent sign of mass of the source (which can still be either positive or negative) and that of the particle experiencing the field (which would always be assumed to be the positive one) then we obtain a framework that is more easily generalizable to a relativistic theory. But it must be clear that the two approaches discussed here are equivalent in the Newtonian context and still require all mass (gravitational and inertial) to be either positive or negative and when

the direction of the gravitational field due to a local matter inhomogene-
ity is not considered to be an observer dependent property we must indeed
consider the equivalent gravitational field to itself be dependent on the sign
of the accelerated mass (which is no longer positive definite), otherwise the
equivalence between the two viewpoints breaks down.

From the viewpoint where the mass experiencing a gravitational field is
considered positive definite, a Newtonian gravitational field experienced by
a particle we would normally consider to have positive mass, if it is not
the result of an accelerated motion far from any matter inhomogeneity (in
which case we would be dealing with an equivalent gravitational field), would
be experienced by a particle we would normally consider to have negative
mass as an oppositely directed Newtonian gravitational field, while the mass
of the particle experiencing this relatively defined gravitational field would
not even show up in the equations used to determine its motion. But if
the gravitational mass experiencing this reversed gravitational field is kept
positive then it must be assumed that the inertial mass is also kept positive
and under such conditions the equivalent gravitational field would appear
not to be reversed. It is because we do not appropriately keep the mass
sign invariant when we try to determine the motion of what we currently
describe as a negative mass particle in a given accelerated frame of reference
that we need to reverse the experienced equivalent gravitational field. But
when the external force applied on what we would currently describe as
a negative mass particle is gravitation itself it is possible to assume that
this force is reversed (from that which would be experienced by what we
currently describe as a positive mass particle), not because the mass of the
particle is reversed, but because the local gravitational field itself is reversed.
In such a case the inertial *force* would not be reversed, because the mass
(both gravitational and inertial) that is experiencing the field is not reversed
and it must also be assumed that the equivalent gravitational field is left
unchanged (from that which is experienced by what we already consider to
be a positive mass particle). Therefore, acceleration still doesn't take place
in the direction opposite the applied force and this is all a consequence of
the fact that even though the local gravitational field appears to be reversed
from such a perspective, the equivalent gravitational field in contrast is left
invariant along with the sign of mass of the particle.

It should be clear, then, that in the context of an approach according to
which the particles experiencing a gravitational field are always assumed to
have a positive mass, the crucial assumption is that while the gravitational

fields attributable to local matter concentrations are dependent on the nature of the body experiencing their effects, the equivalent gravitational field associated with acceleration away from local masses would for its part remain invariant regardless of how the body experiencing it perceives the gravitational fields attributable to local matter inhomogeneities. This hypothesis can be considered to be equivalent to that which in the above described approach consists in assuming that the equivalent gravitational field must actually be reversed for a negative mass, because this is indeed what allows the inertial properties of an object to be independent from its mass sign. I believe that this observation clearly shows that I'm justified in analyzing the problem of negative mass from a conventional perspective according to which the mass experiencing a gravitational field is explicitly assumed to be reversed, because in such a context the underlying assumptions are made more apparent and it is also easier to explain what I'm referring to when discussing the case of abnormally gravitating matter. In a Newtonian context I will therefore continue to use the first viewpoint according to which it is possible for the mass experiencing a gravitational field to be negative.

Now, we may want to dig a little deeper and ask why it is exactly that we are allowed to assume that the direction of the equivalent gravitational field is dependent on the sign of mass of the object experiencing it? I have tried very hard to develop a better understanding of the whole phenomenon of inertia and what I have learned has actually helped me to derive the above discussed results. Indeed, this investigation has enabled me to realize that the assumption that the equivalent gravitational field is reversed when the mass which is subject to acceleration is itself reversed is not just a requirement of the necessary relational definition of the sign of mass, but must be imposed in order to allow a relational description of the phenomenon of inertia itself, in the sense that inertia should be conceived as arising from purely relative motions between matter particles, as suggested by Ernst Mach a long time ago. In this context I have become convinced that the inertial forces acting on a particle can be understood to arise as a consequence of an imbalance, caused by acceleration relative to the global inertial frame of reference (associated with the distribution of matter on the largest scale), in the sum of forces attributable to the interaction of the accelerating particle with each and every other particle in the universe.

What happens in effect is that there must be a similar imbalance of the gravitational forces exerted on similarly accelerating positive and negative mass bodies arising from their interaction with the rest of the matter in the

universe, because the imbalance responsible for the existence of the inertial gravitational force is similar to a skewed mass distribution and if the actual large scale matter distribution responsible for those effects is roughly the same from the viewpoint of both positive and negative masses in the absence of local matter inhomogeneities, then the imbalance should develop in a similar way for both positive and negative masses from the viewpoint of their own mass sign. Thus, what must be retained of this investigation is that the equivalent gravitational field which applies on a negative mass body should in fact be the opposite of that which would be experienced by a positive mass body with the same acceleration that is located within the same matter distribution, even if simply as a consequence of the fact that for a reversed mass the same motion relative to the same matter distribution should give rise to a similar imbalance in the sum of *forces* attributable to interaction with all the matter in the universe.

Indeed, given that the mass itself is reversed, the invariance of this imbalance would mean that the equivalent gravitational field responsible for the inertial force must also be reversed in the accelerated frame of reference, so that the force existing relative to it can itself be left invariant. But if the equivalent gravitational field associated with the acceleration of a negative mass body is the opposite of that associated with the same acceleration of a positive mass body it follows that the reaction to any applied force is indeed the same for opposite mass particles, despite the fact that there is no distinction between inertial and gravitational mass signs (even for negative mass particles). This may be considered to actually explain why it is appropriate to assume that it is the inertial force itself, instead of merely the product of mass and acceleration, that would be opposite the direction of the applied external force for a negative mass body, as the generalization of Newton's second law that I proposed allows to express.

But it must be clear that if there is a requirement for inertial mass to be reversed along with gravitational mass it does not follow from imposing the validity of the equivalence principle as a condition that all matter should have the same motion in the absence of any interaction other than gravitation, as is usually considered. Indeed, as the previous analysis allows to understand, even a negative mass body for which both the gravitational and the inertial masses are negative should not be expected to follow the same trajectory as a positive mass body in the presence of a local positive or negative mass concentration (despite what is usually assumed). What I have tried to explain is precisely that even when inertial mass is assumed to be reversed along with

gravitational mass it is not possible to preserve the validity of the equivalence principle integrally. Thus, a local inertial frame of reference cannot be defined independently from the sign of mass of the body experiencing it given that the direction of the gravitational force resulting from a particular matter distribution depends on the sign of mass of this body. The less restrictive requirement that all matter with the *same mass sign* in the same location follows the same motion (acceleration) is in fact appropriate and restrictive enough for the equivalence between gravitation and acceleration to apply, precisely when it is considered that both gravitational and inertial masses must always be reversed together, because it is only in such a case that at least all positive (inertial) mass matter follows the same motion (it is usually assumed that a negative gravitational mass with positive inertial mass would not), which is all that is really required by the principle of equivalence (masses with the same sign should have the same acceleration) as I have explained. Thus, it is in this particular sense only that we may assume that the equivalence principle requires inertial mass to be reversed along with gravitational mass.

1.5 The equivalence principle with negative mass

It is not usually recognized that the general theory of relativity is actually based on two postulates, because only the first postulate, which concerns the equivalence between acceleration and a Newtonian gravitational field, is well known and is explicitly taken into account. But actually a second postulate is required to obtain the current formulation of the theory and is implicitly assumed to be valid without justification. It is the hypothesis of absolute significance of the sign of energy. This second assumption appears to be necessary in order to preserve the validity of the first postulate under conditions where the presence of negative energy matter would in effect need to be taken into account. But even though the postulate of the absolute definiteness of the sign of energy may be considered problematic in the context of the preceding analysis, it remains to be shown whether it is possible to provide a consistent classical theory of the gravitational field in which only this second postulate would be rejected. Thus, I will try to show in this section and later when discussing the mathematical aspects of a generalized theory

of gravitation that it is perfectly possible and indeed actually necessary to maintain the validity of the equivalence principle in its most essential form while nevertheless rejecting the assumption of an absolute significance of the sign of mass or energy.

Now, I would like to emphasize that the true motivation behind the equivalence principle is to be found in a requirement which we may call the relativity principle and which is actually one particular expression of the requirement of relational definition of all physical quantities. This relativity principle imposes that the state of motion of an object, and in particular its rate of acceleration, is to be determined merely in relation to the state of motion of other physical systems, so that there is no absolute state of acceleration relative to an arbitrarily chosen, unique, metaphysical reference system. The principle that there is an equivalence between a Newtonian gravitational field and an acceleration enables this requirement to be fulfilled, because it allows what might have otherwise appeared to be an acceleration relative to absolute space to merely be a state of rest in the vicinity of a local mass concentration not accelerating relative to the same 'absolute' space, as Einstein understood, but as we tend to ignore nowadays in favor of the mere mathematical requirement of general covariance of the field equations. I think that it must be recognized that, in fact, the only essential implication of the equivalence principle is that indeed there is no longer any motive for arguing that because acceleration is felt (unlike velocity) it must be absolute. Thus, it may appear problematic that even if we find generally covariant equations for the gravitational field in the presence of negative energy matter, the fact that according to the previous analysis such matter would not share the same accelerated motion as positive energy matter in the presence of a local matter inhomogeneity (while it should in the absence of such a perturbation for reasons I explained before), would appear to allow the effects of acceleration relative to matter at large to be distinguished from those attributable to the gravitational field of a local mass.

There is indeed a tension between the principle of relativity and the previously discussed requirements concerning negative mass matter which we may illustrate by once again using Einstein's elevator experiment. Under circumstances where what I have identified as appropriately behaving negative energy matter would be present it may seem in effect that we could differentiate an acceleration of the elevator occurring far from any local mass from an acceleration of the elevator occurring while it is at rest near such a large mass. This is because near a planet or another large matter inhomogeneity

positive and negative mass bodies would accelerate in opposite directions, one toward the local mass and the other away from it (one upward, the other downward), while in the elevator which is simply accelerating far from any large mass, positive and negative energy bodies would share the same acceleration, apparently betraying the fact that the acceleration is 'real'. We may therefore assume that an observer in the elevator would be able to tell when it is that she is simply standing still in the gravitational field of a planet and when it is that she is actually accelerating far from any big mass. The 'true' acceleration would have been revealed to the occupants of the elevator as that for which both the positive and the negative mass bodies have the same acceleration. Consequently, we would seem to be justified to conclude that the notion that acceleration is totally equivalent to a gravitational field (which is the essence of the principle of equivalence) is no longer valid when we introduce negative mass matter with properties otherwise required to make it a consistent concept (according to the preceding analysis).

Indeed, I made it clear before that it is not possible to abandon the principle of inertia or Newton's third law (action and reaction) in order to accommodate the existence of negative mass matter, because if those rules were not strictly obeyed under all conditions then not much else would remain valid. We cannot even tell what a world devoid from this constraint would look like and there is no reason to assume in particular that the equivalence principle itself would still be obeyed, as is usually assumed, because after all this principle is a reflection of the phenomenon of inertia. Trying to save the principle of equivalence by simply allowing negative mass matter to react abnormally to applied forces (as if that was required when inertial mass is negative), so that it accelerates like positive mass matter in the presence of local matter inhomogeneities, would not make sense, because this would mean that the principle of inertia no longer applies in general and again in such a case there is no guarantee that even the alternative situation we expect to observe under those conditions would really occur. I believe that there are reasons why no violations of the principle of inertia have ever been observed despite the fact that the techniques required to reveal such transgressions have long been available. It would not be clever to think that it is by rejecting this principle that we can maintain the requirement of the equivalence between a gravitational field and acceleration. Clearly, there must be something wrong with certain assumptions we take for granted concerning the equivalence principle itself. The fact that this is the principle upon which relativity theory and our modern concept of gravitation is founded should not prevent us from

reexamining some of the implicit assumptions surrounding it. Failing to do so would mean that we have to give up on the idea that negative energy matter could exist, because only so could we then avoid being faced with the annoying and unpredictable consequences of an alternative choice concerning the properties of this matter.

It is important to note at this point that it would be inappropriate to suggest that the requirement that the principle of equivalence also applies in the presence of negative mass matter could perhaps be accommodated if opposite mass bodies were found to always share *opposite* accelerations instead of always sharing the same acceleration as is traditionally assumed. It is certainly true that under such circumstances it would still be impossible to distinguish a true acceleration given that opposite mass bodies would always accelerate in opposite directions, whether those accelerations are the result of the presence of a local concentration of matter or the result of the presence of an equivalent gravitational field far from any large mass. But this situation could only occur if in the context of an appropriate conception of the phenomenon of inertia based on the previously discussed generalized formulation of Newton's second law, the equivalent gravitational field associated with acceleration was not reversed despite the reversal of the mass of the accelerated body experiencing it.

From that viewpoint we should actually expect that one of two opposite mass bodies would fall down while the other would fall up in the accelerating Einstein elevator far from any local mass, even when no force is applied on any of the two masses independently. However, this kind of behavior would constitute an even more severe violation of the principle of inertia than that which would occur in the case of the chasing pair of opposite mass bodies described before, given that in this case there wouldn't even exist any identifiable cause for the upward acceleration of one of the two bodies, because the elevator does not even interact with any of the masses and merely constitutes a reference system. In fact, this situation is so devoid of plausibility that it clearly means that it is not possible to try to salvage the equivalence principle by assuming that the equivalent gravitational field is not reversed for an accelerating negative mass body. The fact that the kind of invariance of the equivalent gravitational field that is involved here would also violate the requirement of relational definition of the sign of mass, as I explained in the previous section, only contributes to confirm the validity of this conclusion. We must therefore accept that while the local inertial frames of reference can differ for positive and negative mass bodies near

some local matter inhomogeneities, they must nevertheless be identical for opposite mass bodies far from local mass concentrations.

I will soon explain why it is exactly that we are allowed to consider that the principle of relativity of motion (concerning acceleration in particular) is not threatened by the conclusion that the free fall state of motion of a negative mass body can be different from that of a positive mass body in the presence of local matter inhomogeneities. But it is important to first point out that in the case of the elevator near a local mass we are in effect considering an inhomogeneous matter distribution for which positive and negative energy matter concentrations are *not superposed* in space (in the classical sense) and therefore do not compensate one another. If such compensations between the effects of *local* matter inhomogeneities were to occur, when for example we would have two superposed gas clouds of opposite energy signs with the same overall motion, possibly rotating, but in the same direction, then the acceleration of positive and negative energy bodies located near or within those matter distributions would have to be the same despite the presence of local inhomogeneities in the configuration of positive and negative energy matter. This actually means that there couldn't be any effect from the motion relative to such a matter distribution, because whatever gravitational effect positive energy matter would have would be compensated by the opposite effect of the similarly distributed negative energy matter present around the body. This is true also of rotation which according to Einstein's theory induces a frame dragging effect which we may assume to be dependent on the sign of mass like any other gravitational phenomenon.

Now, you may recall this earlier discussion (from the preceding section) in which I suggested that it should be possible to attribute the inertial gravitational forces experienced by positive and negative mass bodies in the accelerating elevator away from local masses to some imbalance in the sum of gravitational forces attributable to interaction with all the matter in the universe arising as a consequence of acceleration relative to the reference system associated with the average state of motion of this large scale matter distribution. However, given what I just mentioned regarding the compensating effects of superposed matter distributions with opposite masses and identical motions, it seems that one would have to assume that no imbalance could arise from the gravitational interaction with positive and negative energy matter if they are similarly distributed in space on the largest scale. Thus, one must conclude that if the positive and negative energy matter distributions are indeed mostly identical and are at rest with respect to one another

on such a scale (as appears necessary if the cosmological principle applies equally to both matter distributions), then there should be no effect on both positive and negative mass bodies from the presence of matter on the largest scale.

What this means is that there could not be any imbalance in the equilibrium of gravitational forces attributable to the large scale matter distribution that would give rise to inertial forces or the equivalent gravitational fields, because one imbalance attributable to motion relative to positive energy matter would be compensated by a similar but opposite one arising from the same motion relative to negative energy matter (all masses would experience two opposite equivalent gravitational fields all at once). It thus appears that there is something wrong with one or more of the implicit assumptions entering this deduction, because inertia does exist and indeed if there was no inertia the world would not be anything even remotely similar to what we observe. Of course the idea that there simply is no negative energy matter in the universe (so that the imbalance due to acceleration relative to the positive energy matter distribution is not compensated by an imbalance due to acceleration relative to the superposed negative energy matter distribution) may be tempting, because after all we do not observe any such matter. But keep in mind that it will later be explained that this hypothesis is not required and that in any case it would again amount to simply reject the possibility that such matter may exist without providing any justification for this very convenient hypothesis.

We may summarize the situation by noting that what we know for sure is that if the identical accelerations of the opposite energy bodies relative to the elevator far from any local mass are due to a similar imbalance in the gravitational forces attributable to the interaction of those bodies with matter on the largest scale then this imbalance must be attributed to motion that takes place relative to what are essentially *identical* matter distributions with the same motion and the same rotation and which *should* therefore have opposite effects of equal magnitude on positive and negative energy bodies with the same motion *relative* to this homogeneous matter distribution. If this is recognized, then we have to admit that in the context where negative energy matter actually exists it would be difficult to see how a local inertial frame of reference could be determined by the large scale matter distribution through the gravitational interaction. In such a case it would then seem that we have to conclude that there may need to exist something like absolute acceleration relative to an arbitrarily chosen unique reference system lack-

ing any physical underpinning. What I have understood though (for reasons that will be discussed later) is that the hypothesis that both the large scale positive and negative energy matter distributions have an effect on positive or negative energy bodies considered independently constitutes the incorrect assumption which appears to invalidate the hypothesis that all motion (including accelerated motion) is relative, even in the presence of negative energy matter.

If we drop the assumption that a *negative* energy matter distribution that is uniform on the cosmological scale can exert a force on *positive* energy matter (and vice versa for the effects of positive energy matter on negative energy matter) then it seems that we can explain the imbalance responsible for the force of inertia as being the consequence of an acceleration with respect to the one particular, but relatively defined, reference system which is that relative to which most of the matter in the universe is at rest, because in such a case there would be no canceling of the effects attributable to the positive energy matter distribution by those of the negative energy matter distribution (and vice versa) on the largest scale. Therefore, what I suggest we have to recognize, if only by necessity, is that there is no compensation, for a positive mass body accelerating relative to the average matter distribution on the cosmological scale, between the equivalent gravitational field attributable to positive energy matter and that which we could have attributed to negative energy matter. Similarly, there should be no equivalent gravitational field attributable to acceleration relative to the average distribution of positive energy matter to compensate the equivalent gravitational field attributable to acceleration relative to negative energy matter for a negative mass body. If I'm right this is due merely to the fact that on the cosmological scale particles of one energy sign interact only with the matter distribution that has the same energy sign. I'm particularly confident in the validity of this proposal given that I had actually understood the requirement of absence of interaction between the positive and negative energy matter distributions on the cosmological scale before I even realized that it was required to solve the problem of the relativity of motion in the context where negative energy matter is indeed allowed to exist. I will explain what independently justifies this conclusion in a following section of the current chapter.

What happens, therefore, is that only the very large scale distribution of *positive* energy matter determines the local inertial frame of reference that is experienced by *positive* energy bodies in the absence of local matter inhomogeneities, while only the overall distribution of *negative* energy matter

determines the local inertial frame of reference experienced by *negative* energy bodies in the absence of local matter inhomogeneities (this language would also be appropriate from a general relativistic viewpoint). Thus, what differentiates the situation of the elevator near a large mass of positive *or* negative sign and the situation we have in the elevator accelerating far from any such local mass is that in the first case the force responsible for the observed acceleration is the result of an imbalance that is caused by unequally distributed inhomogeneities in the positive and negative energy matter distributions and this imbalance is dependent on the sign of energy of the body experiencing it (as there are two possibilities for both the sign of mass of the source and that of the accelerated body), while in the latter case the observed force responsible for the acceleration is the result of an imbalance that is always caused by the motion of a body of given mass sign relative to a uniform matter distribution with the same mass sign (necessarily and invariably) so that it is not dependent on the sign of energy or mass of the body experiencing it (positive and negative energy bodies react in the same way to acceleration relative to matter on the largest scales) as long as the distributions of positive and negative energy matter are similar and are not accelerating or rotating relative to one another on the largest scale.

All accelerations are therefore relative accelerations between well-defined physical points of reference within the universe and no absolute state of rest (more exactly of absence of acceleration) can be identified. This is true even if there does exist a unique particular reference system (actually two unique but corresponding reference systems) which is singled out as that relative to which the motion (state of acceleration) of positive and negative mass bodies is the same in the absence of local disturbances, as a result of the correspondence of the average state of motion of the positive and negative energy matter distributions on the largest scales. But this conclusion applies merely in the context where *globally* any particle is gravitationally influenced only by its interaction with matter of the same energy sign whose state of motion relative to the particle therefore alone determines the local inertial frame of reference in which the particle evolves. Thus, despite the exact correspondence of the positive and negative energy matter distributions on the largest scale (which if those sources were locally concentrated would imply an absence of resulting effect on both positive and negative mass bodies) there nevertheless exists a resulting effect from the presence of this matter on a local mass of any sign that allows to determine a unique frame of reference and this is what explains that there appears to be a difference between acceleration far from

any local mass and a gravitational field due to local matter inhomogeneities, while in fact the difference observed is merely the consequence of the fact that a body with a given mass sign interacts only with the large scale matter distribution with the same sign of mass, so that no compensation can exist in this case.

In light of those developments it appears that what the previously discussed insight concerning the nature of the equilibrium involved in determining local inertial frames of references should be understood to mean is that free fall motion, instead of involving a total absence of forces, as is usually assumed in a general relativistic context, must be considered to be a manifestation of the acceleration-dependent equilibrium in the sum of gravitational forces attributable to interaction with both local masses and the large scale matter distribution. This interpretation appears to be required in the context where negative energy matter must be recognized to exist, given that in such a case there cannot even be a unique inertial, or free fall frame of reference dictated by the geometry of spacetime, so that we are forced to consider the reality of the general relativistic gravitational field as being associated with such a physical interaction. Indeed, it is only when we are dealing with a universal force, defined precisely as a force that affects all bodies in the same way, that we can *choose* (as a mere convention) to include this force in our definition of the metric properties of space and time, given that geometry must by definition be shared by all objects in a given space. But this is just a convenient choice which would clearly appear for what it is if the force in question was not affecting all bodies similarly (therefore betraying its material nature).

Einstein himself insisted that if we are to recognize the validity of a principle of general relativity of motion, then the speed of light can no longer be assumed to be constant (even though it is left invariant locally, along a geodesic), given that in the elevator experiment light rays may follow curved paths. But from this viewpoint the curvature of spacetime should naturally be expected to arise as the manifestation of an equilibrium of gravitational forces dependent on acceleration and due to the interaction of the bodies experiencing it with all the matter in the universe (except the large scale matter distribution with opposite mass sign), otherwise it would be impossible to determine what affects the trajectory of light in an accelerated reference system far from any local matter inhomogeneity. Indeed, even in a flat space far from any local matter concentration the motion of light in a straight line, which is usually considered to be a consequence of geometry itself, would

from my viewpoint be a consequence of the equilibrium of forces arising from the gravitational interaction with the rest of matter in the universe. This does not mean, however, that the geometrical interpretation of gravitation is incorrect, but merely that the geometrical properties of space must definitely be conceived as arising from those interactions and more precisely from some sort of equilibrium in the sum of gravitational forces that can be altered by the presence of local matter inhomogeneities. Such a viewpoint has the added benefit of being more easily generalizable to a theory where the gravitational interaction must be described as an interaction mediated by quantum particles.

In any case I think that it is clear that statements to the effect that relativity theory has made the concept of gravitational interaction obsolete and replaced it with that of spacetime curvature (so that gravitation is merely a manifestation of the geometry of spacetime) can no longer be assumed meaningful if curvature is itself a relatively defined property which arises as a consequence of an equilibrium of local and inertial gravitational forces which depend on the sign of energy of the objects involved. I think that the situation we have here is similar to that in which electromagnetic theory was before the quantization of light and the photon concept was proposed, because spacetime is now viewed as a continuous medium, not dependent on underlying physical causes, that directly takes part in determining the motion of objects, just like electromagnetism was originally considered to be a fundamental wavelike phenomenon directly influencing the motion of charged bodies. When it was shown that light is a corpuscular phenomenon the whole notion of electromagnetic wave was not abandoned of course, because there was something real about the wavelike character of electromagnetic phenomena and this is the element which came to be integrated into quantum theory. Similarly, I think that the concept of spacetime curvature cannot and need not be abandoned when gravitation is described as an interaction (which would ultimately be described as a quantum phenomenon) involving some sort of equilibrium which is dependent on the sign of mass of the object submitted to it and from which local inertial frames of reference emerge, only, spacetime curvature can no longer be considered as actually *being* gravitation itself.

As Hans Reichenbach once emphasized [20], if we choose to integrate the gravitational force into our definition of spacetime we may no longer need to explicitly take the force into consideration to explain the motion of bodies, but we must still invoke a force as the cause of the geometry itself.

Thus, it is not gravitation which was replaced by curved geometry, but all of geometry that became a manifestation of the gravitational interaction and I think that this is particularly relevant in the context of a theory of gravitation that allows to take into account the possibility of the existence of negative energy matter. Actually, the commonly made remark to the effect that relativity allowed to eliminate gravitation as a real force appears to be motivated by the fact that the gravitational force arising from local mass concentrations was given the status of inertial force (similar in kind to the Coriolis force) by relativity and given that inertial forces were never seen as real forces, then it is believed that gravitation can now be considered a fictitious force under all circumstances. But I believe that it is rather the contrary that is true and that it is the inertial forces which are allowed by relativity to be considered as real gravitational forces. In such a context the fact that inertial forces are involved in giving rise to the dynamic equilibrium which determines the mass sign-dependent local inertial frames of reference is a further indication that the geometry of spacetime is the product of an equilibrium of real gravitational forces arising from the interaction of any given mass with the rest of matter in the universe.

Having properly identified the origin of the identical response of positive and negative mass bodies to acceleration, I do not want to immediately enter into a discussion as to what are the true elements of justification behind the assumption that particles of one mass sign are not affected from a gravitational viewpoint by the presence of matter of opposite mass sign on a cosmological scale. But it may nevertheless already be noted that the fact that one particular reference system appears to be singled out as having unique status among all possible states of acceleration is not a unique feature of the approach described here. Actually, in a general relativistic context, even in the absence of negative energy matter, this feature of our description of the motion of objects should appear all the more natural given that all inertial frames of reference are an outcome of the gravitational interaction and are therefore determined by the surrounding matter distribution. There exists in effect one very particular reference system in our universe which we may call the global inertial frame of reference and which is that which is determined by the average motion of all masses together and relative to which most masses in the universe do not accelerate (or rotate). That there may be such a unique point of reference does not mean that it is not relationally defined. Relativity theory allows to explain the existence of this particular reference

system as being a result of the combined gravitational interactions of a local body in any state of motion with all the other masses in the universe (with the same mass sign) and therefore in relation to the average motion of those masses. Indeed, even far from any big mass there remains the gravitational effect of the universe as a whole, which can never be ignored. Thus, the situation we usually refer to as corresponding to an absence of gravitational field and which we expect to be experienced far from any local mass is not different in fact from that occurring in the presence of a local mass, only it is characterized by the fact that the gravitational field is then attributable to uniform distributions of either positive or negative mass matter, which incidentally implies coinciding inertial frames of reference for positive and negative mass bodies.

The fact that inertial frames of reference are always determined by the average state of motion of matter in the universe becomes particularly obvious when we consider the reference system associated with a felt motion of rotation which, as experiments have revealed, must be one that is in rotation relative to the most distant galaxies and therefore relative to the largest visible ensemble of matter in the universe. The reference system relative to which a positive mass observer feels no rotation must then be determined simply by the gravitational field attributable to all matter particles with the same mass sign present in the universe in a way that is dependent on the average state of motion of those particles and as such is definitely unique even though its description involves only relationally defined properties. We may still consider the average matter distribution on the largest scale to be rotating, but then its gravitational field would give rise to a rotating inertial frame of reference which, through relativistic frame dragging, would put the whole matter content of the universe in rotation with it[3]. Since Einstein there is no longer any mystery with the existence of such a preferred frame of reference and what I'm trying to explain is that there is also no problem with the fact that there is a unique reference system relative to which at once positive and negative mass bodies have no acceleration when free from

---

[3]It has been mentioned that a (positive mass) observer uniformly rotating with respect to the distant stars and which would choose to consider himself motionless would observe a gravitational field which from a Newtonian viewpoint could not exist, therefore weakening the equivalence principle. But it is interesting to observe that this difficulty would no longer exist in the context where a repulsive gravitational field that grows in proportion to the distance from an axis could be produced by an appropriately configured inhomogeneous, static distribution of negative energy matter.

external non-gravitational forces. We are not faced here with a metaphysical reference system associated with absolute acceleration, but merely with an ordinary reference system relative to which the effects of the gravitational interaction of local masses with matter on the largest scale imposes an absence of acceleration for both positive and negative mass bodies.

Again it must be stressed that even when it may seem that we are dealing with empty space, what the objects actually experience are the effects of the whole surrounding matter distribution conveyed by the gravitational field as an intermediary material entity, which in a general relativistic context actually determines the possibly distinct local inertial frames of reference affecting positive and negative energy bodies. This aspect of the general relativistic (or physical) space is what allows to conceive of rotation as being purely relative, even when the distance of some objects to the rotation axis of a rotating observer becomes large enough that the objects would actually have to move at faster than light velocities in the reference system tied to the observer. Indeed, it is the rotation of the whole gravitational field, as a material entity (which would also occur in a universe totally devoid of 'real' matter), that explains that this motion of the remote objects is possible as a true motion, because locally the objects are not moving (accelerating) relative to the gravitational field (or the local inertial frames of reference), which is then itself rotating, and this is what makes their large velocities and accelerations possible, as is already well understood.

But if acceleration occurs merely relative to the inertial reference systems determined by the gravitational field it must not be forgotten that the state of motion of matter also contributes to determine the gravitational field and therefore it should naturally be expected that there is no acceleration of matter as a whole relative to the global inertial frame of reference determined by the gravitational field produced by this large scale matter distribution. It may also be remarked that the situation we are dealing with here concerning the relativity of acceleration in the presence of negative energy matter is similar to that regarding the relativity of velocity, because there also exists a preferred frame of reference relative to which the temperature of the cosmic microwave background is mostly uniform and which may appear to define a state of absolute rest, but this unique frame of reference is merely that which is not moving relative to the average state of motion (not acceleration) of matter on the largest scale. If there is no conflict with the principle of relativity in such a case, then there need not be a problem in the case of the global inertial frame of reference singled out as being that relative to

which there is no difference between the states of acceleration of freely falling positive and negative mass bodies.

There would then be no substance to the argument that the apparent distinction between acceleration and gravitation which appears to be revealed by the distinct motions of positive and negative energy bodies in the standing still elevator near a local mass allows absolute acceleration (or absolute absence of acceleration) to be determined. Indeed, the local gravitational fields and the associated local inertial frames of reference are always determined in a relative fashion as dependent on the presence of the local masses which are the source of the fields, while the reference system where the states of acceleration of positive and negative energy bodies are identical is determined as that relative to which the large scale matter distribution (which we may assume to be unique to positive and negative energy matter) is itself not accelerating. This all follows from the fact that positive and negative energy bodies interact only with the homogeneous matter distribution with the same sign of energy as their own on the cosmological scale[4], so that motions relative to those matter distributions must be treated differently from motions relative to local matter inhomogeneities, although they are still relative motions.

It must be noted, however, that if the homogeneous large scale distributions of positive and negative energy matter were in motion relative to one another there would then actually be two different global inertial frames of reference associated with the two types of mass (positive and negative) experiencing them, even away from any local mass. In such a case it would be more difficult to differentiate between the situation of the elevator far from any large mass and that in which unequally distributed concentrations of positive and negative mass matter are present locally. It remains, though, that if positive and negative energy matter are produced together during the first instants of the Big Bang as a result of energy conserving creation out of nothing processes (as I will propose in section 3.5 based on results that will be discussed in section 1.9) then we should not expect negative energy matter to be accelerating or even only moving on the average (on the largest scale)

---

[4]In fact, as I will later explain, the large scale distribution of negative energy matter may exert an influence on positive energy bodies, but only when inhomogeneities are present in this matter distribution. The nature of those interactions is such, however, that there is necessarily a cancellation in the sum of the effects involved on the largest scale, so that there can be no overall effect and the same is true for the effects of positive energy matter on negative energy bodies.

relative to positive energy matter in the primordial universe and therefore also at the present time, to a certain extent.

Based on the above discussed considerations I have thus come to the conclusion that, after all, the principle of relativity is not really threatened by the introduction of negative energy matter obeying the requirement of relational definition of its mass sign. But clearly the equivalence principle itself (which allows accelerated motion to be treated relativistically) is no longer to be considered valid in the sense it was traditionally believed to be and if it need not and indeed cannot be abandoned it must, however, be generalized or somewhat relativized. In fact, we already know for sure that the equivalence principle always applies only in local frames of reference whose states of motion can be different in various locations. We can tell in effect that a gravitational field is attributable to the presence of local masses instead of being the consequence of an acceleration, even in the total absence of negative energy matter, when we consider a portion of space that is sufficiently large. For example, if we consider two elevators suspended on opposite sides of a planet, instead of a single elevator, it is obvious that even though observers in each of those elevators could assume that they are accelerating far from any local mass, from the global viewpoint where we would be observing oppositely directed gravitational fields and an absence of relative motion of the elevators we would have to conclude that those fields are due to the presence of a local mass and not to acceleration relative to the homogeneous large scale matter distribution, even in the absence of negative mass bodies in the elevators. In fact, even in a single elevator standing still on the surface of a small planet, freely falling positive mass particles would have a tendency to slightly converge toward one another, therefore betraying the fact that the observed acceleration is an effect of the presence of a nearby mass attracting the particles toward its center. Yet we do not consider the equivalence principle to be violated under such conditions.

What I'm suggesting therefore is that instead of assuming that the equivalence of gravitation and acceleration applies only locally, we have to recognize that it really applies only for a single elementary particle, which would be the most localized physical system we may consider. If we assume that no two such particles can be exactly superposed in an elementary volume of space (which ultimately may be true for bosons just as for fermions if there is a maximum local energy level associated with the Planck scale) we could say that the hypothesis that the equivalence of acceleration and gravitation applies merely within a local free fall frame of reference is equivalent to the

assumption that the equivalence principle applies only for a single elementary particle at once. But then such a particle could have either positive or negative mass and the equivalence principle could be considered to apply not merely to one particle at once, but to one particle with one mass or energy sign at once, which would be a simple generalization of the discussed hypothesis and as such should not raise any further issue (of the kind I have considered so far). For one elementary particle with one energy sign there would never be a difference between acceleration and a gravitational field. It is only when we consider two or more particles of *any* mass sign together, or more precisely in relation to one another, in the presence of a gravitational field attributable to a local matter inhomogeneity (when there is no compensation between the gravitational fields attributable to the local positive and negative energy matter distributions) that we can tell the difference between acceleration relative to the large scale matter distribution and such a gravitational field, but this may be assumed irrelevant when we are considering that no two particles (especially two opposite mass particles) can actually be found in the exact same position at the same time.

It is generally recognized, however, that what makes gravitation different from other interactions is the fact that the motion of bodies in a gravitational field does not depend on the physical properties of those bodies (when no other force field is present). But even though this characteristic would appear to be violated in the presence of negative energy matter obeying the consistency conditions I have identified, this does not make gravitation any less distinct. Indeed, in the context of the previously discussed viewpoint where it is the direction of the gravitational field attributable to a given matter distribution which varies upon a reversal of the mass of the particle submitted to it (which would actually be considered positive definite), the equivalence principle would merely be relativized by the presence of such negative energy matter, because the difference between the motion of positive energy bodies and that of negative energy bodies would actually be a consequence of the different measures of spacetime curvature which (as I will explain later) can be associated with those two measures of the Newtonian gravitational field. But in such a situation it appears natural to expect that opposite mass bodies should not be restricted to share the same local inertial frames of reference, because in fact they do not even evolve in the same space, but in spaces characterized by different metric properties.

Thus, the fact that the gravitational field can be conceived in such an observer dependent way means that in the case of gravitation it is not the

reaction that varies when the 'charge' is reversed, but the field itself, so that it would still be true that, in any given situation, all bodies (sharing the same measure of the gravitational field) follow the same motion (acceleration does not depend on the detailed characteristics of the bodies experiencing the same gravitational field). The equivalence principle can thus be assumed to still be valid in the presence of negative energy matter, only it would apply separately for positive and negative energy bodies (just as it applies separately for separate portions of space), because each of those two kinds of matter particle is to be attributed its own free fall frame of reference defined in relation to its mass sign. Therefore, all particles with the same energy sign, whether their energy is positive or negative, would still share the same local inertial frame of reference and this is all that is truly required for a general relativistic gravitational field theory to apply.

1.6   An effect of voids in the matter distribution

It is sometimes recognized that there is a kind of equivalence between the presence of a void in an otherwise uniform matter distribution and what would be the presumed effect of the presence of gravitationally repelling matter present in a quantity and with a distribution equivalent to that of the missing matter. In the context of an expanding universe we would indeed observe underdense regions of the cosmos to be producing a local acceleration of the rate of expansion, while overdense regions would produce a local deceleration of it. The acceleration observed in the case of underdense regions would have all the characteristics of a gravitational repulsion originating from those regions, which would force the matter still remaining inside their volume to migrate to the periphery of what would become the observed voids in the matter distribution [21]. The same effect would also cause nearby underdense regions to merge into even larger spherical voids, as if they were attracted to one another by the force of gravity. This is what all authors who have considered the issue agree must occur when underdense regions form in an expanding universe. Thus, in this particular case, it seems that the gravitationally repelling matter formations would actually be submitted to mutual gravitational attraction with similar formations, even while they would repel oppositely configured formations consisting of overdense regions

and would presumably also be repelled by them.

But it is usually considered that there is nothing more than an accidental analogy between the case of those matter formations and any gravitationally repulsive matter, because if the effect occurs as described above then, according to the traditional understanding, such gravitationally repulsive voids would have to have not only negative gravitational mass, but also positive inertial mass [22] and as everyone 'knows' this kind of negative mass is forbidden by the equivalence principle and relativity theory, which require the equality of gravitational and inertial masses. Thus, what we would *observe* to be happening is not what most people would consider should occur if we are actually dealing with gravitationally repulsive matter. Indeed, as I previously explained what is usually assumed is that gravitational repulsion is a kind of definite and invariable property of matter of some type and that this kind of matter would therefore itself also be repelled by matter of the same type. This is usually assumed to be the unavoidable consequence of attributing a negative inertial mass to negative energy matter. But, given the previous discussion and the insights I provided concerning what should be a consistent concept of negative mass or negative energy matter, it should be clear that we would not be justified to argue that the observed phenomenon involving voids in a uniform matter distribution does not replicate the behavior we should expect of negative mass matter. In fact, from my viewpoint it rather seems that the described interaction between overdense and underdense regions of an expanding universe would be exactly that which we should expect to occur if positive and negative masses were actually involved. Therefore, we cannot so easily reject the possibility that the discussed phenomenon is actually telling us something important about the nature of negative energy matter.

I do believe that there is actually more than a valid analogy between voids in a uniform positive energy matter distribution and gravitationally repulsive matter and that there is something very profound which we need to understand concerning the phenomenon described here. Indeed, I think that the discussed equivalence should not be restricted to the case of expanding matter, but must be considered valid even in a local context where the rate of universal expansion is a negligible factor. But if the gravitational dynamics of voids in a homogeneous positive energy matter distribution actually reflects that which we should expect of a phenomenon involving gravitationally repulsive negative energy matter then it may suggest an interpretation of negative energy matter which would have to do with an absence of pos-

itive energy of some kind. It must first be explained, however, why it is that we may actually be allowed to consider that the equivalence discussed above is valid exactly and constitutes a very general feature of the gravitational interaction despite the objections which might be raised against that possibility.

Basically, what we may object concerning the idea that the presence of a void in a uniform positive energy matter distribution could be equivalent to the presence of an excess of negative energy matter is that it is usually assumed that there can actually be no net gravitational force inside a spherical void in a uniform matter distribution that would be attributable to matter outside the void, a conclusion that seems to be supported by Birkhoff's theorem [23]. What Birkhoff's theorem implies is that there can be no net gravitational force from matter outside any spherically symmetric region in a uniform matter distribution that may itself be considered to be spherically symmetric. This is usually assumed to imply that there cannot be any net gravitational force inside a spherical void in a uniform matter distribution given that such a matter distribution is in effect homogeneous and isotropic. This assumption would actually mean that in the absence of any matter inside a spherical region there can be no gravitational force at the boundary of the region, as any acceleration could only be attributable to matter inside the region considered and there would then be no matter inside that region.

The influence of voids on the local rate of acceleration of cosmic expansion which was discussed above would thus merely be a result of the fact that the rate of growth of the distance between two galaxies located on the boundary of such a void actually depends on the density of matter *inside* the void and given that this density would be lower than the average then the rate of growth of the distance, or the local rate of expansion would be larger in proportion with the amount of matter missing inside the void. But that does not mean that it is usually assumed that there would actually be a repulsive gravitational field on the surface of the void. In fact there appears to be some confusion surrounding the issue discussed here, as some authors recognize that there cannot be an equilibrium of gravitational forces in the presence of a void in the cosmic matter distribution and yet they fail to recognize that this may actually give rise to repulsive gravitational fields for the surrounding positive energy matter, probably because they assume that the effect of the noted disequilibrium would be that which is observed to affect the local rate of expansion, while actually this is a distinct (but not entirely unrelated) effect associated merely with cosmic expansion. But what I believe must

be recognized is that there would in effect be gravitational repulsion in the presence of an underdensity in an otherwise uniform matter distribution, not only at the boundary of the surface, but everywhere inside the void with a net force that would decrease to reach a null value as we approach the center of the void. This situation would then clearly be different from that we would have in the case of a hollow sphere of finite size inside of which the Newtonian gravitational field should indeed be zero everywhere.

It must in effect be understood that contrarily to what is usually believed, Birkhoff's theorem does not forbid this conclusion, because the decisive condition entering this theorem is that of spherical symmetry, which would actually be obeyed if we were considering a hollow sphere or a universe that was spherically symmetric around *any* point on any scale, but which I suggest would fail locally for a universe with an actual void in its matter distribution. Indeed, the case of a homogeneous and isotropic universe is equivalent to that of a sphere of finite size only when the universe is considered at the scale at which its matter is uniformly distributed and no significant void is present, which explains why Birkhoff's theorem (which is a necessary element of current cosmological models) is observed to apply on a cosmological scale. But I think that it would only be in the case of a spherical region centered on an actual sphere of matter of finite size located within an otherwise empty universe that the theorem discussed here would actually remain valid regardless of the distribution of matter inside the spherical region, because only in such a case would we be dealing with a spherical symmetry that is not dependent on the position of the observer. What we usually fail to recognize is that the fact that the matter distribution in the universe would be symmetric around *any* location in the absence of a void in its homogeneous and isotropic matter distribution means that the presence of a void would necessarily alter the equilibrium of forces around that void.

It is clear indeed that in the presence of a uniform matter distribution extending throughout the universe an equilibrium exists locally between the sum of forces attributable to the interaction of a freely falling body with all the matter in the universe and therefore the removal of a certain quantity of matter in a region of finite volume must have an effect that would be the opposite of that which we would attribute to the presence of an equivalent *additional* quantity of matter in the same region of the same universe (in the absence of the void). This should be expected to occur due to the fact that the removal of a certain amount of positive energy matter to create a void would eliminate the attractive gravitational force which would otherwise be exerted

on positive energy matter by the matter in the void and given that there was
no net force before the creation of the void then the other forces which were
initially present must now give rise to an acceleration directed away from the
void and of similar magnitude to that which would have been produced by
the matter that filled the void. Thus, for positive energy matter there would
appear to be a repulsive gravitational force originating from the presence of
a void in such a uniform matter distribution, which would actually be the
consequence of an uncompensated gravitational attraction attributable to
the positive energy matter outside the void. But this is a valid conclusion
only when we recognize that Birkhoff's theorem is not valid in the sense it is
usually assumed to be and that the case of a spherical distribution of matter
of finite size with a central cavity is *not* equivalent to the case of a void in a
uniform cosmic matter distribution.

What must be understood is that if, in the case of a hollow sphere of finite
size, the subtraction of matter to create the cavity does not result in a net
force originating from the matter surrounding the cavity *that is part of the
sphere* this does not mean that it would also be the case that there would be
no acceleration inside the cavity resulting from the gravitational interaction
with the *all* the matter that is present in the universe (unless it was actually
assumed that the universe is empty except for the presence of the sphere).
What is wrong therefore is the idea that when we are considering a spherical
region of the universe the rest of the universe surrounding that region can be
considered as a hollow sphere simply on the basis of the fact that according
to the cosmological principle matter is distributed uniformly in all directions.
In fact, such a spherical region in a uniform matter distribution would be
free of uncompensated external forces only if it was itself filled with matter
as uniformly distributed as the matter found outside the region (which is
actually verified on a cosmological scale in our universe), because it is only
in such a case that the spherical symmetry would apply to any point inside
the spherical region. Again, it must be noted that, in this context, the fact
that the concept of the hollow sphere is nevertheless appropriate to describe
the dynamics of the universe on the largest scale is due merely to the fact
that we do not actually consider the case where spherical voids are present
in the matter distribution, but really the case of a uniformly filled matter
distribution for which no spherical regions devoid of matter are present on
the particular scale that is considered (as a requirement of the cosmological
principle).

It must be clear that I'm not suggesting that there would be uncompen-

sated gravitational forces in the case of the finite size hollow sphere itself (if it was located in an empty universe for example). In fact, the problem here has to do again with the fact that we fail to apply the requirement of relational definition of physical properties when we are dealing with the resultant effect of the gravitational forces attributable to the universe as a whole. Indeed, from the traditional viewpoint, when we are dealing with a chosen spherical region of the universe we are implicitly assuming that the surrounding matter which may influence the particles located inside that region (through the gravitational interaction, even if there is no net force) is spherically distributed around the center of the spherical region considered, as if the location of the center of mass of the universe was an intrinsic invariable feature of the whole configuration. But the center of a matter distribution in a physical universe without boundary is not an absolute feature (as would be the case for a hollow sphere), but must be defined in a relational manner as any other property, if we are to be able to determine the consequences on a given object of being located in such a position. When we are dealing with the matter distribution in a universe without spatial boundary and in which the local inertial frames of reference are determined by the entire matter distribution (following Mach's principle) the true center of mass defined in terms of the influences exerted on a given body is *always* located right at the position where that body is to be found, wherever this position may be in the matter distribution.

Thus, a particle located *at the center* of a void in a uniform matter distribution could actually be considered to be in the situation of a particle in a hollow sphere, because for this particle the whole sphere of influence of the universe is centered on the void (in this situation the surrounding matter actually is a hollow sphere centered on the particle's position). Therefore, such a particle would feel no uncompensated gravitational force from the whole universe, as required. But if this particle moves to one side or another in the void, the matter distribution influencing the particle in its new position would be centered on the new position and this means that the void in the previous hollow sphere is shifted to the opposite side, just as the sphere itself is shifted in the direction of the particle's new position. The symmetry of the initial configuration would therefore no longer be present and the equilibrium of forces would no longer apply. In the new configuration a whole layer of matter must be 'removed' on one side of the external surface of the imaginary hollow sphere (in the direction opposite the particle's displacement) and added on the other (this is easier to visualize in a closed universe)

which, given the distances involved, means that an enormous amount of matter has changed position from the viewpoint of the particle. It must therefore be recognized that in the final configuration the void in the imaginary sphere is no longer centered on the center of mass of the sphere, but is actually located away from the center of the sphere. As a consequence, the spherical symmetry from which depended the conclusion that there would be no net gravitational force inside the sphere is no longer to be found in the final configuration experienced by the particle and therefore it must be expected that there would be a net gravitational force on the particle and an acceleration relative to the matter distribution.

It is important to understand that however large you consider the imaginary sphere encompassing the matter distribution (the size of the universe) to be when dealing with the effects of the gravitational interaction with the whole universe, if the center of the sphere is shifted to one side there would be a non-negligible effect from the displacement of its center of mass. This is true even if the distance to the periphery of the sphere (where the changes occur) is very large and the strength of the gravitational interaction decreases with the square of the distance, because the larger the distances (the larger the sphere) considered, the larger the quantity of matter that is shifted from one side to the other and thus the larger the changes involved in the local gravitational field. We should not be surprised, then, that even the retarded interaction with matter so distant could have an effect similar in magnitude to the effect that would be exerted by the matter missing from a void located near some particle experiencing those forces. If the center of mass of the universe is always located at the position of the particle experiencing the gravitational effects of all the infinitesimal elements of matter in this universe, then the local effect of the absence of gravitational attraction from those portions of matter which would be present if a nearby void in the positive energy matter distribution was absent would necessarily result in a net force on positive energy matter arising from the gravitational attraction of all portions of matter located on the opposite side of the void. But such a force would be completely equivalent to a repulsive gravitational force arising from the void itself.

The fact that from a practical viewpoint the formation of a local void in a uniform positive energy matter distribution would actually have to occur through the expulsion of positive energy matter outside the region that is to become the void and therefore would necessarily produce a compensating overdensity of negative energy matter in the region surrounding the void

would not forbid the existence of a net repulsive force on positive energy matter inside the void, even though it does in effect mean that there would be no resulting force on matter located some distance away from the void. If we consider for example the ideal situation of a spherical void produced through the creation of a surrounding spherical shell of positive energy matter at higher than average density, then as long as a positive energy particle is located outside this shell it would feel no net force, because any reduction of attractive force from the void would be compensated by an increased attractive force arising from the presence of the shell. But as soon as the particle would enter the shell it would begin to experience the equivalent gravitational repulsion, because the outer layers of the shell would no longer provide any net force on the particle while the void for its part would still exert its net effect, because the equivalent repulsive force it produces is attributable to *all* the surrounding matter (whose distribution is centered on the position of the particle) and not just to the spherical shell. Thus, the case of the particle which experiences no gravitational force *at the center* of a void in a uniform matter distribution is merely a particular case of the more general description where there is actually a net force everywhere inside the void, except at the exact location of its center, as would be the case if we were considering the gravitational attraction attributable to an isolated sphere full of matter (like a planet or a spherical gas cloud). This is an important result which will have decisive consequences for a consistent description of the nature and properties of negative energy matter.

Concerning the insight just described it is important to note that even if under certain circumstances there may be an equivalence between an imbalance in the sum of gravitational attractions attributable to all the positive energy matter elements in the universe and what would appear to be a gravitational repulsion exerted on a positive energy body, we are nevertheless always dealing with gravitational attraction. Indeed, there is no question that it is the gravitational attraction of positive energy matter that is responsible for the apparent gravitational repulsion which would be exerted on a positive energy body by a void in the otherwise uniform positive energy matter distribution. It is clearly as a consequence of the fact that positive energy matter is missing in the direction where the void is located, while the matter present in the opposite direction still exerts its gravitational pull, that there exists a net force directed away from the void.

Thus, what looks like a gravitational repulsion exerted in a given direc-

tion by some matter configuration and which could from a certain viewpoint be equivalent to it would actually be the product of a gravitational attraction arising from an absence of matter exerting a compensating attraction in the opposite direction. This is particularly significant in the context where local inertial frames of reference are to be considered as always arising from a perturbation of the equilibrium of inertial gravitational forces by the gravitational forces attributable to local matter concentrations, as I have emphasized in the preceding section. Yet the fact that we are here dealing only with gravitational attraction does not rule out the validity of the analogy which may exist from a classical viewpoint between the presence of true gravitationally repulsive, negative energy matter and an absence of positive energy of some sort. In fact, it rather seems that what allows an interpretation of negative energy matter as being equivalent to an absence of positive energy to be valid as a general feature of gravitation theory is the possibility that always exists (not only in the case of voids in a uniform matter distribution) of attributing an apparent gravitational repulsion to uncompensated gravitational attraction.

To explain what motivates that conclusion I may recall the previous discussion concerning the occurrence of negative energy in certain experiments described using the methods of quantum field theory. There I pointed out that the absence of some positive energy states from the vacuum in certain limited regions of space (between the plates of two parallel mirrors for example) can actually give rise to a vacuum with negative energy density in the volume considered, because removing positive energy from a vacuum state whose energy is already minimum is like decreasing the energy below its zero point into negative territory. The fact that the vacuum is known to have only a very small energy density should not be considered an obstacle to the occurrence of large negative energies in such a way, because as I will explain later in this chapter and in section 3.2 this small energy density appears to be the outcome of very large (actually maximum) but (mostly) compensating opposite energy contributions, which could be reduced to an arbitrarily large extent by the conditions which are responsible for locally decreasing (under particular circumstances) the energy of the vacuum below the equilibrium point. But if we may, in effect, attribute a negative energy to some configurations in which particular states are missing from the vacuum along with their contribution to the total energy of this vacuum, then there is no reason why we could not consider that negative energy states in general are equivalent in some ways to an absence of positive energy from the vacuum,

if from a phenomenological viewpoint there is no distinction between those two situations.

I must again mention in this regard that many authors have expressed doubts concerning the concept of vacuum energy as arising from fluctuations involving virtual particles and have suggested that there may be nothing real with the processes so described outside of the context where they are occurring as part of otherwise real processes involving 'real' particles. But I think that it is precisely the fact that the existence of those processes would imply the reality of negative energy states that really motivates this mistrust, because it is no secret that for most physicists the theoretical possibility of the existence of negative energy states is not well viewed. However, I believe that this aversion is merely a consequence of the fact that the traditional concept of negative energy matter is in effect not viable and that it has not yet been realized that a better description of negative energy matter is possible and even necessary, as I emphasized before.

In any case, the idea that virtual processes would only occur as part of otherwise real processes, thus explaining why we must nevertheless consider the effects of such fluctuations when calculating transition probabilities, is meaningless, because in a given universe anything that occurs is related (directly or indirectly) to everything else and even in empty space, far from any 'real' matter, the virtual processes of particle creation and annihilation characteristic of the quantum vacuum would occur as an integral part of the surrounding real processes to which they are causally related as a consequence of their common origin in the Big Bang. In fact, I will explain in section 3.9 why those considerations actually constitute a decisive element of a consistent cosmological theory, even aside from the issue of vacuum energy. Therefore, the argument that the negative energy states predicted to occur in the vacuum under the right conditions are not real, because our description of the vacuum is itself not appropriate in general, cannot be retained. Also, the fact that it has been confirmed that the cosmological constant is not absolutely null is a strong motive to conclude that the rejection of the reality of vacuum fluctuations as essential aspects of our description of empty space is not vindicated from the viewpoint of observations and therefore that negative energy states are a real possibility.

I have already explained why we should expect to observe mutual gravitational attraction between two bodies with the same sign of energy and gravitational repulsion between opposite energy bodies. But on the basis of my conclusion concerning the nature of the gravitational force on a positive

energy body that would be attributable to voids in a uniform positive energy
matter distribution we now also have the possibility to assert what would be
the effects of missing positive energy from the vacuum. Indeed, given that
the vacuum is to be conceived as involving a constant and uniform density of
energy on the largest scale, any negative local variation in its density must
share the features of voids in a uniform matter distribution. It therefore
appears that if the presence of voids in an otherwise homogeneous positive
energy matter distribution does in effect produce an equivalent gravitational
repulsion on positive energy bodies, then the absence of positive vacuum
energy in localized regions should actually exert an equivalent gravitational
repulsion on the surrounding positive energy matter. This would occur as a
result of the fact that an absence of positive energy from a region of the vac-
uum would result in an uncompensated gravitational attraction from the sur-
rounding positive energy vacuum pulling positive energy matter *away* from
the region where the energy is missing. From that viewpoint we can thus
deduce that the physical properties (related to the gravitational interaction)
that we should expect to be associated with missing positive vacuum energy
are the same properties which I explained we should expect to be associated
with the presence of negative action matter, which confirms that from a phe-
nomenological viewpoint negative energy matter is gravitationally equivalent
to an absence of positive energy from the vacuum.

Given this equivalence between negative energy and absence of positive
energy from the vacuum, it follows that if states of negative vacuum energy
are allowed by current theories then we must conclude that negative energy
matter is itself allowed to exist and may not always be constrained by the
limitations observed to apply in the currently considered experiments where it
occurs merely as a consequence of the suppression of positive energy from the
vacuum, attributable to singular configurations of otherwise positive energy
matter. It must be recognized, however, that if the presence of negative
energy matter in a region of space is equivalent for positive energy matter
to an absence of positive energy from the vacuum this is simply because
in general for an equilibrium state of any kind the presence of a negative
contribution is equivalent to the absence of a positive contribution of the
same magnitude and it just happens that the vacuum is a physical system
that appears to arise from precisely such an equilibrium state. But we must
remember that a void in a uniform matter distribution of a given energy
sign (not involving the vacuum) is physically different from a local absence
of vacuum energy of the same sign, even if in both of those cases the effects

are equivalent to the presence of an excess of matter of opposite energy sign, because in the first case we are dealing with an *absence* of matter of a given energy sign, while in the latter case we are actually dealing with the *presence* of matter (of opposite energy sign).

At this point it is important to mention that there would occur a phenomenon of gravitational repulsion similar to that described above, but which would apply from the viewpoint of negative energy matter in the presence of voids in a negative energy matter distribution or in the negative energy portion of the vacuum. Indeed, using the same logic that allowed me to derive the consequences of the presence of a void in a uniform positive energy distribution it is possible to deduce that the absence of negative energy from an otherwise homogeneous matter distribution would actually be equivalent from a gravitational viewpoint to the presence of a concentration of positive energy matter. One assumption that will be crucial for my derivation of the modified general relativistic gravitational field equations is indeed that the equivalence described here is valid both ways and that positive energy matter can always be considered to actually consists of voids in the *negative* energy portion of the vacuum, which makes the whole situation symmetrical in a way that does not even depend on the viewpoint of the observer. It must be clear, however, that I'm not suggesting that positive energy matter is equivalent to voids in a filled distribution of negative energy matter, even if I do suggest that we must assume that an absence of negative energy matter from an otherwise uniform distribution of such matter would indeed have gravitational effects similar to those attributable to the presence of positive energy matter. I must emphasize once again that a void in a uniform *matter* distribution remains clearly distinct from a void in the uniform energy distribution of the *vacuum*. This means that my proposal is distinct from Dirac's failed hole theory (proposed as an attempt to solve the negative energy problem), in particular because what I'm suggesting is that *all* positive energy matter particles (and not just antimatter particles) are actually equivalent to voids in the negative energy portion of the vacuum rather than in a filled continuum of negative energy matter.

What Dirac proposed in effect is that all negative energy states are already occupied, so that positive energy fermions at least should not be expected to make transitions to those negative energy states. But even if the existence of such a filled, uniform continuum of negative energy matter was to have no effect on positive energy matter (perhaps due to its uniformity), the fact that

from my viewpoint there would be no reason to assume that positive energy states are not completely filled in the same way means that this hypothesis would not agree with observations. Indeed, it is not possible to assume, in a theory that respects the requirement of a purely relational definition of the sign of energy, that positive energy antiparticles are merely voids in a completely filled negative energy matter continuum, as Dirac proposed, without also assuming that negative energy antiparticles would be voids in a completely filled positive energy matter continuum. But, given that positive energy states are obviously not all occupied by matter particles, it appears that this requirement cannot be satisfied. We may then instead assume that *all* positive energy particles are voids in a filled negative energy matter continuum, but again in such a case we would have no reason not to assume that all negative energy particles are also voids in a filled positive energy matter continuum. The problem, however, is that it seems impossible to assume that we could have a completely filled distribution of negative energy matter and at the same time a completely filled distribution of positive energy matter if negative energy matter is to also consists of voids in a filled distribution of positive energy matter, because so many voids in the positive energy matter distribution as would be necessary to describe the filled negative energy matter distribution would leave no possibility for the positive energy matter distribution to itself be nearly completely filled.

What cannot be assumed therefore is that negative energy states are completely filled and positive energy particles are voids in this negative energy distribution while positive energy states are completely filled and negative energy particles are voids in this positive energy distribution, because those two possibilities are mutually exclusive (cannot occur together). But while it may perhaps appear appropriate from an observational viewpoint to assume that we simply have a filled negative energy matter continuum combined with a nearly empty distribution of positive energy matter, there would also be problems with such a proposal. Indeed, what reason would we have not to assume that it is only the positive energy matter distribution that is filled (even though this assumption would clearly contradict observations)? The problem is that we cannot in effect postulate that both positive and negative energy matter are voids in their respective opposite energy matter distributions if we also postulate that there is no absolute (non-relational) difference between positive and negative energy matter. In other words, it is not possible to assume symmetry under exchange of positive and negative energy particles if matter of a given energy sign is to be conceived as voids in the

matter distribution of opposite energy sign and this simply because matter cannot be at once present and absent. The truth is that any description of matter or antimatter as voids in a matter distribution of opposite energy sign would require giving preferred status to negative energy matter as being the matter whose distribution is completely filled (because obviously the positive energy matter distribution at least is not completely filled) and this would break the requirement that only differences in the energy sign of particles are to be conceived as physically significant.

What must be clear, therefore, is that if we were to make use of such a description we would allow the identification of a preferred sign of energy as being that which would be associated with the filled matter distribution, while from a theoretical viewpoint that should be considered impossible. A theory of matter as voids in a uniform, opposite energy matter distribution would in effect imply that the requirement of symmetry under exchange of positive and negative energy matter is violated in a way that cannot be allowed if the sign of energy is to be conceived as a relationally defined physical property. Thus, it must be recognized as forbidden to consider that the presence of matter with a given energy sign could be explained as resulting from the presence of voids in a *matter* distribution of opposite energy sign, even if there does exist a phenomenological equivalence between the effects of missing positive or negative vacuum energy and the absence of matter from a homogeneous distribution with the same sign of energy, because again those are two distinct phenomena.

The contradiction which would occur if we were to assume that positive energy particles are voids in a filled uniform distribution of negative energy matter, while negative energy particles are voids in a filled uniform distribution of positive energy matter is that we would require the presence of a lot of particles of both energy signs to fill the matter distributions and at the same time the presence of a limited number of particles of both energy signs due to the presence of all the voids attributable to the presence of the nearly filled opposite energy matter distributions. According to my proposal by contrast it becomes possible for both positive and negative energy particles to actually exist as real observable particles independently from the presence of one another. Thus, if the voids in the negative energy portion of the vacuum, which I assume to be equivalent to the presence of positive energy matter, are not equivalent to voids in a hypothetical filled distribution of negative energy matter it is simply because in fact voids in the vacuum cannot be equivalent to an absence of voids in the vacuum.

I may add that from the viewpoint of a consistent interpretation of negative energy matter there would also be a problem with Dirac's original proposal that a void in the filled negative energy continuum could be created along with a positive energy particle (as would a particle-antiparticle pair) when photons provide enough energy to raise a negative energy particle to a positive energy level. Indeed, as I mentioned before and for reasons I will explain in section 1.8, a consistent theory of negative energy matter would require that negative energy matter be dark, which means that there would be no electromagnetic interactions between opposite energy particles and therefore a positive energy photon could not even interact with a negative energy electron to provide it with the required positive energy. Thus, even if we insist on assuming the existence of a filled negative energy continuum we could not use this hypothesis to explain the existence of antimatter.

It is essential to understand, therefore, that the situation we would have if all negative energy states were filled is different from that we would have when dealing with a vacuum in which there would be a very large negative contribution to the energy density of zero-point fluctuations. Indeed, in contrast with the vacuum, a negative energy matter distribution which would be filled at one particular epoch would no longer be filled at a later time given that space is expanding. This is reflected in the fact that vacuum energy obeys an equation of state which is different from that of a homogeneous matter distribution. Also, even if there is a large negative contribution to the energy of the fluctuating vacuum there is no reason to expect that it gives rise to a situation similar to that which would occur if space was filled with negative energy matter, because in such a case there must also be a large positive contribution to the energy of empty space (the motives behind this conclusion will be clarified in section 3.2). A space filled with positive or negative energy matter would be as different from the true vacuum as the primordial soup which existed in the first instants of the Big Bang is different from the space nearly devoid of particles that currently exists between galaxies. Thus, if a theory of voids is to have any relevance in a gravitational context it must involve a description of matter of any energy sign as consisting of voids in the opposite energy portion of the vacuum, so that the presence of matter with a given energy sign does not imply an absence of *matter* with opposite energy sign.

When the energy distribution in which the voids equivalent to the presence of positive energy matter occur is the negative energy portion of the *vacuum* it therefore becomes possible to assume the presence of arbitrarily

high or arbitrarily low densities of matter of both energy signs all at once in the same region of space, because in effect the presence of matter of one energy sign in a given location does not preclude the presence of matter with an opposite energy sign in the same location (at least when the matter distributions are smooth enough). Thus, we do not need to assume the presence at all times of a nearly filled negative energy matter continuum combined with a distribution of positive energy matter of arbitrarily low density, which would otherwise be the only (perhaps) observationally acceptable configuration, but which would also have allowed to establish an absolute (non-relational) distinction between positive and negative energy matter, as I just explained. But what makes the vacuum particularly suitable for accommodating the above proposed description of matter as consisting of voids in some uniform energy distribution is the fact that we are actually allowed to assume that there are both positive and negative contributions to vacuum energy density, even as arise from otherwise identical virtual particles. We can therefore expect a certain level of compensation between the gravitational effects of those two contributions that may give rise to an arbitrarily small residual value for the cosmological constant. Indeed, in sections 3.2 and 3.5 I will explain that one of the consequences of the assumption that there exists a distinct component to the energy of the vacuum arising from the presence of those virtual particles that directly interact only with negative energy matter is that the natural value of the cosmological constant which we can expect to observe is zero, even though this value can be altered so as to compensate any imbalance that might have existed during the first instants of the Big Bang between the densities of positive and negative matter energy, in the limits imposed by the weak anthropic principle. This is an important result which will have an impact on many aspects of cosmology theory.

What is perhaps even more significant, however, is that when we understand that all positive and negative energy particles are actually equivalent to voids in their respective opposite energy portions of the *vacuum*, as I propose, then the unsatisfactory categorical distinction between matter and vacuum becomes meaningless. This is because in such a context all matter can actually be considered to consist in a particular manifestation of some property of the vacuum. It is by building on this insight that I will be able to provide a unified and totally symmetric description of the gravitational dynamics of positive and negative energy matter according to which the measure of energy of matter is significant merely in relation to an energy scale associated with objective properties of the vacuum. I was able to obtain

those results only at a relatively late stage of my reflection, because I had initially assumed that only the nearly vanishing total energy density of the vacuum could have an influence on matter of any energy sign and that the positive and negative contributions to vacuum energy could not be considered independently from one another. But once I realize the inappropriateness of this hypothesis, the above discussed results emerged as clearly unavoidable and extremely significant. The notion that both positive and negative energy particles are actually voids in their respective opposite energy portions of the vacuum therefore appears to be the ultimate embodiment of the requirement of a relational definition of all physical properties understood as a basic consistency condition that must apply to any physical theory.

Concerning the effects which I'm suggesting should be attributed to energy missing either from a homogeneous matter distribution or from the homogeneous vacuum we may ask to what extent a void may actually be considered as physically significant in the sense of being merely an anomaly in an otherwise uniform distribution of matter or energy. If we examine the situation carefully it becomes clear in effect that given that for both matter and vacuum it must be the surrounding energy that exerts the *outward* directed gravitational pull that would be experienced as a gravitational repulsion, then it follows that as we consider voids of larger sizes there may come a point when there would be no matter left outside the void to produce the uncompensated attraction that must exist to produce the equivalent repulsion. Normally this is not an issue, as any void that forms in a matter distribution which can be assumed to be arbitrarily smooth initially (and this appears to be a necessary feature of our universe at the Big Bang as I will explain in chapter 3) will necessarily involve the creation of a surplus of matter in its surroundings, which for a remote observer would have the effect of compensating the equivalent force arising from the presence of the void itself, as I previously mentioned. Such voids, regardless of how large they may become, would therefore leave the universe at large in a state equivalent to that of a uniform matter distribution, which would allow it to continue to exert its influence in the empty regions.

But if we are to consider the equivalence between missing positive vacuum energy and the presence of negative energy matter to be generally valid, then the presence of a uniform negative energy matter distribution would imply the existence of a void in the positive energy portion of the vacuum which would actually extend to the whole universe. This void would have been

present in the vacuum from the very beginning of the universe's history and would not have developed through the production of some inhomogeneity. In such a case we would no longer be able to assume the existence of an uncompensated gravitational pull on positive energy bodies from the surrounding positive vacuum energy, because indeed there would be no surrounding vacuum energy with higher positive density to effect the attraction. Under such conditions, therefore, I'm allowed to conclude that no outward directed gravitational force which we could assimilate with an equivalent gravitational repulsion would exist.

Now, given that I will later argue that the equivalent gravitational repulsion exerted on positive energy matter by voids in the positive energy portion of the vacuum actually constitutes the only form of gravitational interaction between this matter and negative energy matter, it would appear that the preceding conclusion imposes very strong limitations on such an interaction. Indeed, it transpires that the absence of equivalent gravitational repulsion on positive energy matter from a completely homogeneous negative energy matter distribution, is a very general and unavoidable feature of the description of the gravitational interaction between positive and negative energy matter. This is because such a limitation would also be verified in the case of a uniform distribution of positive energy matter from the viewpoint of negative energy bodies if the gravitational repulsion exerted on those objects by positive energy matter can be attributed to an absence of *negative* energy from the vacuum.

Thus, if opposite energy bodies can be shown to interact only through their respective vacuums, we would be allowed to conclude that negative energy matter interacts with positive energy matter only in the presence of inhomogeneities in any of the two matter distributions. But given that only an inhomogeneity that develops over the initially smooth negative energy matter distribution (if we may suppose that negative energy matter is as homogeneously distributed as positive energy matter on the cosmic scale) can contribute to the gravitational dynamics of positive energy matter and given that the formation of such an inhomogeneity would involve the formation of a compensating one involving an opposite variation of density in the surroundings of the first, we must then conclude that the presence of an average density of negative energy matter has absolutely no effect (at least from a gravitational viewpoint) on the gravitational dynamics of positive action matter (and vice versa). This would mean in particular that the rate of universal expansion of positive energy matter cannot be influenced by the

presence of negative energy matter and similarly that the expansion of negative energy matter is not affected by the presence of positive energy matter. This, again, is a very significant result whose implications will be developed in chapter 3.

I may add that the conclusion discussed here is the one on which is founded the hypothesis discussed in section 1.5 which allowed a relational description of the phenomenon of inertia. There I explained that if both the large scale positive and negative energy matter distributions were to exert an influence on positive energy bodies, then the hypothesis that accelerated motion is relative would be invalidated in the presence of negative energy matter on a cosmological scale. Indeed, under such circumstances there would be equal and opposite imbalances in the sum of gravitational forces (to which we would try to attribute the resultant inertial force) arising from the acceleration of a positive mass body relative to the two opposite energy matter distributions whose average states of motion should correspond with one another on the largest scale. But if only matter of positive energy has a gravitational effect on positive energy bodies on the cosmological scale, then the global inertial frame of reference experienced by a positive energy body could actually be determined by the average state of motion of positive energy matter given that the inertial force exerted on such a body would result only from its gravitational interaction with the large scale distribution of positive energy matter. Thus, we can now see why the rejection of the assumption that a uniform negative energy matter distribution can exert a force on positive energy matter (and vice versa), which appears to be required for a relational explanation of the phenomenon of inertia based on the principle of relativity, was in effect justified. The preceding discussion actually shows (when we recognize that positive and negative energy matter can interact only through the effect they exert on the opposite energy portions of the vacuum) that this hypothesis is not only desirable, but actually constitutes an unavoidable consequence of the description of negative energy matter as being equivalent to missing positive vacuum energy.

But in the context where the description of negative energy matter as being equivalent to voids in the positive energy portion of the vacuum is similarly applied to positive energy matter (in the sense that positive energy matter would be equivalent to the presence of voids in the negative energy portion of the vacuum) a further distinction would arise. Indeed, just as negative energy matter would interact with itself independently from the fact that it is equivalent to voids in the positive energy portion of the vacuum,

positive energy matter, as voids in the negative energy portion of the vacuum, would have to still interact with itself, which means that such voids do interact with themselves. In fact, even if the missing negative energy was uniformly distributed throughout all of space it would still exert an influence on itself despite the fact that a similar distribution of missing positive vacuum energy would have no effect on *positive* energy matter, that is, on voids in the *negative* energy portion of the vacuum. In other words, the fact that a void in the negative energy portion of the vacuum, which is equivalent to the presence of positive energy matter, could leave no outside surrounding negative energy to affect the behavior of negative energy matter (if this void is uniformly distributed throughout the entire volume of the universe) would not affect the ability for such a void to gravitationally attract *positive* energy matter or other voids in the negative energy portion of the vacuum, because in such a case the interaction is actually occurring between the matter particles themselves (or the voids) and not between a particle and the surrounding vacuum with the same energy sign.

Finally, it may be of interest to mention that if we were to consider the effect on a positive energy body of a void in a uniform negative energy *matter* distribution then based on the above discussed insights we should deduce that the outcome would be a gravitational attraction directed toward the center of the void. This could be predicted to occur in two different ways. First, given that we can now expect negative energy matter to exert a gravitational repulsion on positive energy bodies, then on the basis of what has been learned concerning the effects of voids in a uniform matter distribution we could conclude that the absence of gravitational *repulsion* in the direction of the void consequent to the absence of negative energy matter in this void would give rise to an uncompensated *repulsive* force directed toward the center of the void, which would be equivalent to a gravitational attraction directed toward the center of that same void, but which would actually arise from the gravitational repulsion of the surrounding negative energy matter. But given that we now also know that a uniform distribution of negative energy matter has no influence on positive energy bodies it would seem preferable to derive the consequences of an absence of such negative energy matter based on an alternative approach which borrows from the results discussed in the preceding paragraphs.

Indeed, what allows me to conclude that a uniform negative energy matter distribution has no effect on positive energy bodies is that the presence of

such uniformly distributed matter is equivalent to a void of universal proportion in the positive energy portion of the vacuum which therefore leaves no surrounding positive energy to produce uncompensated gravitational forces. But then, if you remove negative energy matter in a portion of this void the resulting configuration would be that of an imperfect void or an imperfect distribution of *absence* of positive energy from the vacuum. But a local absence of *absence* of energy is really just the same as a local presence of energy and if the energy that was absent (when negative energy was present) was positive then the energy that is locally present will itself be positive. This local absence of negative energy matter will thus be totally equivalent to the presence of an equal amount of positive energy matter and should therefore be expected to produce on positive energy bodies a gravitational attraction directed toward the void. This is an effect which may have interesting consequences on the cosmological scale, in the context where variations in the density of negative energy matter would have a magnitude comparable with the average density of the matter itself. I will explore the practical consequences of this important result in section 3.3. But for now let me mention that the effectiveness of the preceding description is a further confirmation of the existence of a close relationship between vacuum energy and matter energy, while the high level of symmetry involved also indicates that the description of the properties of negative energy matter proposed above fully agrees with the requirement of a relational definition of the physical attribute of energy sign.

1.7 Six problems for negative energy matter

The preceding discussion may already make us feel more comfortable with the possibility that there could exist negative energy matter, despite the traditional reluctance to accept the reality of negative energy states. But at the current stage of my account this confidence would not yet be totally appropriate. Even in the context of the new understanding unveiled in the previous sections there indeed remain many problems associated with the possibility that negative energy matter may exist in our universe. First of all, we do not observe in the universe any matter or celestial object which would clearly appear to be involved in repulsive gravitational interaction with other material bodies. This is a very basic but also very constraining fact. Associated with this problem is the fact that the current predictions

of quantum field theory are based on a systematic rejection of the possibility of a transition to negative action states (as states of negative energy propagated forward in time) and yet they appear to produce results which agree very well with observations in all situations where the nature of the interactions involved is well understood and appropriate use of the associated computational methods can be made. This could provide an additional motive for arguing against the possibility of the existence of negative energy matter. Such pieces of evidence certainly cannot be dismissed without very good reasons. Any theory involving particles propagating negative energies forward in time must explain why it is that we can safely ignore the existence of those particles in formulating a quantum theory of elementary particles and their interactions, even while we would presumably have to take their effects into account in a classical astronomical context where the effects of the gravitational interaction are not negligible.

A second category of difficulty has to do with the possibility that seems to be allowed, in the context where negative energy particles would exist, for the annihilation of particle-antiparticle pairs to occur in which one of the particles would have negative action, therefore permitting matter to vanish, leaving absolutely nothing behind. This would of course require the annihilating opposite energy particles to also have opposite electric and other non-gravitational charges, because charge must still be conserved. We have no reason, however, to assume that negative action matter does not also come in two varieties, one propagating negative energy and all non-gravitational charges forward in time and the other propagating positive energy and the same charges backward in time (so that we have opposite charges from the forward time viewpoint). Therefore, we cannot a priori reject the possibility that such annihilations could take place. But that is a much worse problem than may perhaps appear to be, because if such annihilations were possible there would then be no reason why the time-reverse processes could not also take place. If that was the case it would actually mean that pairs of opposite action particles could be spontaneously created out of nothing without immediately returning to the vacuum like ordinary particle-antiparticle pairs given that the process could occur without requiring a violation of energy conservation.

The fact that opposite energy particles would gravitationally repel one another should not prevent an annihilation process involving such opposite action pairs from taking place, as the gravitational interaction is very weak and the fluctuations present in the vacuum could still allow the process to

occur at least when charged particles are involved, because the opposite charges carried by the particles would give rise to attractive forces that would counter the gravitational repulsion. Indeed, if the electrostatic attraction between opposite charges does not prevent ordinary particle-antiparticle pair creation processes from occurring then there is no reason why such an effect would need to be taken into account in the case of pair annihilation processes involving opposite action particles. In any case the fact that the gravitational repulsion between opposite energy particles would not affect the possibility for the associated *creation* processes to occur means that the problem is real. It may, therefore, seem like positive energy matter particles could annihilate to nothing at an arbitrarily large rate upon encounter with negative energy particles, or else be created out of nothing abundantly even long after the Big Bang, while both kinds of phenomena would clearly violate observational constraints which actually provide no evidence at all that such events are taking place. This category of difficulties may then appropriately be called the energy out of nothing problem.

A third potential problem has to do with the possibility that appears to be offered as a consequence of the existence of negative energy states for ordinary positive energy matter particles or even any preexisting negative energy matter particles to 'fall' into the allowed negative energy states in a continuous unstoppable process during which they would either release positive radiation or absorb negative energy radiation and reach ever 'lower' energies. This is a difficulty which would also affect negative energy matter as it is traditionally conceived and which is known as the vacuum decay problem. It would arise from the fact that the zero energy level would no longer constitute a minimum level of energy (the ground state) at which there could no longer be any transition to lower energies. Here we appear to have a situation where the existence of negative energy states raises the specter of allowing an arbitrarily large amount of work to be generated out of nearly nothing (by letting matter fall into the negative energy states and using the energy difference to produce work), as if energy conservation alone was not enough to restrict the evolution to negative energy states. This is clearly another issue of incompatibility with observation, because such decays are not observed to occur, even under the previously discussed conditions where negative energy densities are allowed to occur in a limited way by ordinary quantum field theory. In this context we may in effect ask what it is that prevents positive energy particles from falling into the lower negative energy levels which are predicted to exist under particular circumstances by quantum

field theory? This is all by itself a legitimate question which has remained unanswered. Even from the viewpoint of the traditional interpretation of negative energy states this situation looks like a deep mystery.

But what is probably the most difficult problem which one must face upon recognizing the necessity of introducing a notion of negative energy matter obeying the requirements of a relational definition of physical quantities (which imply that opposite energy bodies must gravitationally repel one another) is that the existence of such matter may appear to allow violations of the principle of conservation of energy. This issue arises as a consequence of the fact that it seems possible for energy and momentum to be exchanged between positive and negative energy systems in a way that is similar to that by which positive energy systems exchange energy among themselves. Basically, it appears that the positive energy of a positive energy body can be turned into an equal amount of negative energy belonging to a negative energy body and vice versa when a 'collision' between two such opposite energy bodies would occur. For example, positive energy could be lost by a positive energy body colliding with a negative energy body initially at rest, while negative energy would be gained by the negative energy body with which the first body has interacted (or vice versa). This would give rise to a net variation in the total energy of the two bodies that would be equal to twice the individual change of energy (rather than allowing a cancellation of changes, as is observed when two positive energy bodies collide). The solution to that problem will have to arise from a proper understanding of some remarkable consequences of the insights gained while solving the first category of problems discussed above.

A further difficulty could arise in the context where the inertial force on a negative mass body has the same direction as that which applies on a similarly accelerating positive mass body, despite the reversal of inertial mass which I have argued must occur when gravitational mass itself reverses. Indeed, from the viewpoint of an improved conception of the phenomenon of inertia based on a generalized formulation of Newton's second law it is no longer possible to consider that acceleration would take place in the direction opposite the applied force for a negative mass body and given that the equivalent gravitational field due to acceleration would be reversed for such an object it follows that the inertial force it would experience is identical to that which is experienced by a similar positive mass body. It would therefore appear that while the presence of a negative mass body could contribute to reduce the gravitational mass in a region of space in which positive mass

matter is also present, it would still provide the same resistance to acceleration despite the fact that it would also provide a negative contribution to the inertial mass contained in this volume. This may not be a problem when we are dealing with independent physical systems with opposite masses but, as I previously mentioned, when a bound system is involved the energy contained in the field of interaction between its constituent particles would be opposite that of the system as a whole and in such a case it would seem that while the energy of the field should reduce the gravitational mass of the system it should nevertheless contribute to increase its resistance to acceleration. Given that bound systems with various force field configurations are quite common, it would seem that objects made of different materials should experience distinct accelerations when submitted to a gravitational force, but no such variations are observed. Some much needed clarification is required here if the concept of negative mass which I have proposed is to be considered viable from an observation viewpoint.

One last potential category of arguments which one might believe could disprove the validity of the idea of gravitationally repulsive, negative energy matter does not actually have to do with the concept of negative energy matter developed here, but merely with more traditional concepts of 'antigravity' and gravitational repulsion. The problems involved would be difficulties for a theory according to which ordinary antimatter is gravitationally repulsive. They would also constitute a challenge for the traditionally favored description of negative energy matter according to which gravitational repulsion is an absolute property of negative energy itself, while gravitational attraction is an absolute property of positive energy matter (so that negative energy matter repels positive energy matter and is attracted to it). If such conceptions where to be retained as valid they would allow paradoxical situations such as perpetual motion and time travel to arise. Given that for most people those difficulties are associated with the general concept of negative energy it is important to explain why the issues involved here would not affect a more consistent theory of gravitationally repulsive, negative energy matter such as that which will emerge from the developments I introduced in the preceding sections.

We are then faced with six categories of problems which appear to undermine a conception of physical reality according to which matter would be allowed to occupy energy levels below zero. I have wrestled with the questions raised by those difficulties for a long time and on many occasions I had nearly given up on the possibility to ever be able to find appropriate

answers that would perhaps explain why negative energy is not an inappropriate concept for physical theory. But, gradually, I came to understand that the problems really have to do with some incorrect implicit assumptions we make when considering the expected behavior of matter in a context where those negative energy states are actually allowed. In the next six sections I will explain the nature of the insights required to appropriately deal with those severe problems.

1.8   The origin of repulsive gravitational forces

When as a young man I first started to contemplate the possibility that there could exist matter in a state of negative energy I soon realized that if such matter was to attract matter of the same type while it would repel ordinary matter and be repelled by it (as I had intuitively assumed should occur, ignorant of the dominant paradigm), then this matter would have to be dark, because nowhere was it mentioned that we observe gravitational repulsion arising from the presence of any planet, star or galaxy. While I was working on improving my understanding of physics in general and trying to develop a theory incorporating the concept of negative mass I simply assumed that negative mass particles where such that they would interact with ordinary matter only through gravitation. I remember that I had read that Feynman once said that we must not question *why* things are the way they are, but simply try to describe in the most accurate way possible *how* they behave. Thus, for a while I was comfortable with the idea that negative energy matter simply does not interact other than through the gravitational force with ordinary matter (although it could interact with itself through the whole spectrum of forces), even if I had no idea why that should be the case and had to assume that this is just the way things are. The only concern I had regarding this situation is that it appeared odd that negative energy matter should not interact with ordinary positive energy matter through the same interactions by which positive energy particles were interacting among themselves, given that negative energy matter could be assumed to actually be composed of the exact same particles as positive energy matter. But then came the shock.

I had for some time tried to figure out what determined the repulsive or attractive nature of an interaction which clearly depends on the signs of the charges of the interacting particles and had slowly came to realize that this

property seemed to be related to the sign of energy of the field of interaction, not yet fully aware that it was actually rather the attractive or repulsive nature of an interaction (determined by the sign of the charges involved) that determined the sign of energy of the field and not the opposite. In any case I had understood that the energy of a field associated with a repulsive interaction between positive energy particles, for example the energy of the electromagnetic field between two electrons, is always positive, while the energy of a field associated with an attractive interaction between positive energy particles, for example the energy of the electromagnetic field between an electron and a positron, is always negative. But it also had to be the case (as I will explain below) that the energy of a field associated with a repulsive interaction between negative energy particles is always *negative*, while the energy of a field associated with an attractive interaction between negative energy particles is always *positive*. What that means is that when two negative action particles are attracted toward one another or bound together in a single system, the contribution of the attractive field mediating the interaction to the energy of the whole system should be positive, while for positive action particles it would be negative.

As I was trying to make sense of this observation in the context where the interaction involved would be that between a positive and a negative energy body I suddenly realized that a catastrophe had just happened. Indeed, if this relation between the sign of energy of the field and the attractive or repulsive nature of the related interaction was right in general it meant that if there was any gravitational interaction between positive and negative energy bodies it should be either repulsive for positive energy matter and attractive for negative energy matter (if the field was attributed positive energy) or repulsive for negative energy matter and attractive for positive energy matter (if the field was attributed negative energy), but never repulsive for both the positive and the negative energy bodies involved in the interaction. This is because a repulsive field would have to have positive energy for a positive energy matter particle, while this same positive energy field would have to exert an attractive force from the viewpoint of a negative energy matter particle for which the same relation would exist in general between the *difference* between the signs of energy of the matter particle and its field on the one hand and the repulsive or attractive nature of the associated interaction on the other (the problem is not restricted to gravitation). This is again a consequence of the requirement of relational definition of the physical properties associated with attraction and repulsion which cannot be considered to be

determined by the energy sign of the interaction field only, but must be a consequence of the difference between the energy sign of the field and that of the matter particles submitted to the force associated with this field.

But it was just nonsense to conclude that an interaction could be both attractive and repulsive at the same time and it is even more so now, in the context where we must recognize that the hypothesis of the mutual gravitational repulsion between positive and negative energy matter is also required for a relational description of the gravitational interaction between those two types of objects. The conclusion I had to draw was thus very clear: no definite energy sign could be attributed to the fields of interaction between positive and negative energy particles (as must be the case for any interaction involving particles with the same sign of energy) and therefore there simply cannot be any interaction between those two types of particle, not even gravitational. This appeared to be a fatal blow, because if there are no interactions of any kind between positive and negative energy matter, then how could negative energy matter have any relevance to the world we experience?

When I realized the existence of this difficulty for a theory of negative energy matter I had already come to appreciate the many advantages that there would be if such matter was allowed to exist (if it could indeed gravitationally interact with ordinary matter). This is because I had been able to solve important problems using even the incomplete description I had by then managed to develop and it seemed improbable to me that the whole idea could simply be wrong. I know that this may look like it was more a hopeful wish than a rational conclusion, but in fact it was actually both hope and reason. Indeed, we had struggled with the problems I was able to solve for a very long time and there really appeared to be no viable alternative solutions to those problems, while theoretically the basic idea of negative energy had a lot of appeal. It is as a consequence of the fact that I had so much confidence in the validity of the basic concept of a symmetry between positive and negative energy states that I did not stopped working on developing the idea when I encountered the difficulties discussed here. And as it turned out the problems encountered became just another challenge on the way to a satisfactory solution to the problem of negative energy.

So, I went from having to explain why there would be no electromagnetic interactions between positive and negative action matter to having to explain why there can be any interaction at all between the same two kinds of matter. Of course I was glad that at least I now had an explanation for why there

is indeed no electromagnetic or other non-gravitational interactions between opposite energy particles, because it was clear that on the basis of the above discussed observations it had to be recognized that there cannot be any direct quantized interactions (mediated through the exchange of interaction bosons) between such particles. But gravitation is different, because it is not yet described as a quantized field and I had hope that it might be its singular classical character that would allow the existence of *some kind* of interaction. It must be clear, however, that the problem described above is very real and unavoidable and its significance should not be underestimated as it actually means that there can be no interactions and no exerted force *between positive and negative action particles*. It must also be understood that this is not a hypothesis, as no consistent theory could describe the interaction of positive and negative mass particles and this must simply be taken as an indication that such interactions are in effect nonexistent.

At this stage you may remember that when I explained that there must be an equivalence (for a positive energy body) between the effects arising from the presence of a void in a uniform positive energy matter distribution and those which we may identify with a gravitational repulsion directed away from the void, I insisted that this repulsion was really the consequence of an uncompensated gravitational *attraction* directed away from the void. Therefore, when dealing with matter distributions which are uniform on a cosmic scale we can observe gravitational repulsion to arise from what are actually purely attractive gravitational interactions. I also insisted that negative energy matter would be equivalent from a classical gravitational viewpoint to the presence of missing positive energy from the vacuum, while the vacuum can itself be considered as being equivalent to some extent (only in this respect) with a uniform matter distribution. But this means that the gravitational repulsion experienced by a positive energy body and which we would expect to arise from the presence of negative energy matter actually results from an uncompensated gravitational attraction attributable to the surrounding positive energy portion of the vacuum. In other words, we can explain the gravitational repulsion apparently exerted by negative energy matter as really consisting of a gravitational attraction involving only positive energy sources.

Thus, even if we assume an absence of direct interaction between positive and negative energy bodies we can nevertheless expect to obtain an equivalent repulsive gravitational force between these objects. It is in this particular sense that the concept of gravitationally repulsive matter developed here can

indeed be assumed to involve effects which are analogous to the situation we have in the case of voids in a uniform matter distribution. But under such circumstances the above discussed problem of the impossibility of direct interactions of either gravitational or non-gravitational kind between positive and negative energy particles is turned into an advantage, because it actually forbids any interactions to occur between opposite energy particles except for the equivalent gravitational repulsion just described and this is precisely what we need. It must be clear, in effect, that the conclusion that there should exist indirect interactions between opposite action particles only applies to gravitation, because an absence of energy from the vacuum does not correlate with an absence of charge, which would be a distinct phenomenon (opposite action particles are not necessarily opposite charge particles)[5].

Those results should be encouraging, as the category of problems they allow to solve was the most basic and the most serious of those which I identified above as facing a theory of negative energy matter. Thus, it is now possible to explain why it is that we have never observed gravitationally repulsive matter, because indeed such matter, if it exists, should not be visible, as it would not interact with ordinary positive energy matter through the long range electromagnetic interaction. It is also possible to explain why it is that the predictions of quantum field theory made under the hypothesis that negative energy states are not allowed in the formalism produce very accurate results which correspond with observations to a very high degree of precision. Because if, in effect, only the equivalent repulsive gravitational interaction just described exists as a kind of influence of negative energy matter on the processes involving positive energy particles which are described by quantum field theory, then given the weakness of the gravitational interaction there should only be a marginal impact from the existence of this negative energy matter on estimations of physical observables currently made under the assumption that negative energy particles do not exist. Indeed, if we do not need to take into account the effects of the attractive gravitational interaction between ordinary *positive energy* matter particles in such calculations, then we should certainly not expect to have to take into account any effects from the equivalent repulsive gravitational interaction with the very sparse amount of negative energy particles that could perhaps be found to wander

---

[5]I am myself guilty of having once assumed, for no serious reason, that an absence of energy would necessarily imply an absence of charge, before realizing that this idea would be highly problematic while remaining completely unjustified from both a theoretical and an observational viewpoint.

around apparatuses located on Earth. Thus, if I'm right, we would have here the solutions to two quite serious problems which were never addressed by any of the authors that previously discussed the possibility of gravitationally repulsive matter, because it can now be understood at once why gravitationally repulsive matter is dark and why it nevertheless interacts gravitationally with ordinary matter.

It must be noted however that even in the context where we have to assume that there is no direct interaction between positive and negative energy particles it would be wrong to consider that positive energy matter interacts only with the positive energy portion of the *vacuum* and not with the negative energy portion of it, because, as I explained in section 1.6, positive energy matter must itself be assumed to consist of voids in the negative energy portion of the vacuum and as such certainly cannot be considered to behave independently from this negative energy vacuum. Yet it should be clear that we are not really dealing with an interaction between opposite energy particles here, but merely with the gravitational interaction of this negative energy portion of the vacuum with itself. Such a phenomenon is somewhat similar to the gravitational dynamics of a uniform negative energy matter distribution in which voids may also be present that would exert attractive gravitational forces on each other and repulsive forces on the rest of the negative energy matter. In such a case it is clear indeed that even if we could assimilate the voids with the presence of positive energy matter, their effects would actually be the outcome of the interaction of negative energy particles among themselves. We may therefore still consider that there is no *direct* interaction of any kind between positive and negative energy matter or vacuum, but again this does not mean that positive energy matter does not experience the gravitational effects of the negative energy portion of the vacuum or that negative energy matter does not experience the gravitational effects of the positive energy portion of the vacuum, because if positive energy matter is a manifestation of negative vacuum energy it cannot be expected that this portion of the vacuum does not interact with itself and the same can be said of negative energy matter as a manifestation of positive vacuum energy. This conclusion will obviously have enormous consequences for the description of the cosmological effects of vacuum energy that will be discussed in chapter 3.

Finally, I may add that a further justification for the fact that we do not yet have strong evidence for the existence of negative energy matter is that, given that such matter is submitted to gravitational repulsion by ordinary

matter and is also gravitationally attracted to itself, it should be expected to migrate away from concentrations of positive energy matter and to concentrate itself in regions of the universe where there is a lesser density of positive energy matter. It would therefore be difficult to observe anomalous gravitational effects which could arise from the presence of celestial objects composed of gravitationally repulsive matter in a region of the universe like ours, where positive energy matter can be assumed to be the dominant form of matter given its relatively large density. In fact, at this point, the lack of evidence for negative action matter has been so well justified that it appears that if we are to ever obtain direct confirmation for the existence of this matter it will be necessary to use alternative methods of investigation and to concentrate on the ability which may be offered in this context to predict features of the visible very large scale matter distribution with better accuracy than current models which neglect the effects of this invisible gravitationally repulsive matter distribution. I will discuss the opportunities that may arise for making such decisive observations in sections 3.3 and 3.4.

1.9   No energy out of nothing

Before we can conclude that there should indeed be no interference with current predictions made using quantum field theory from allowing the existence of negative energy particles in stable states we must first explain why it is that there should be no creation or annihilation processes involving pairs of opposite energy particles with opposite charges, as such a phenomenon could also disrupt current predictions. This is the second category of problems I previously identified as potentially affecting the viability of the negative energy matter hypothesis. Given the plausibility of the hypothesis that negative energy particles should be very rare in our region of the universe it may seem that the problem of the annihilation of opposite energy particles does not constitute a decisive issue. But, as I previously mentioned, we cannot avoid having to face the related problem of the creation of pairs of opposite energy particles, because in such a case it would appear that no favorable initial conditions are required for the discussed processes to occur. Thus, an explanation must be provided for why matter is not, under normal conditions, being created out of the vacuum in massive amounts, despite the fact that the processes involved can occur without violating the principle of conservation of energy, because this prediction clearly disagrees with obser-

vations which indicate a complete absence of such processes, at least under ordinary circumstances.

One may perhaps suggest that given that the opposite energy particles emerging from a creation event in opposite directions would have their momenta both pointing in the same direction (because we must assume that a negative action particle would have momentum opposite the direction of its velocity) could prevent the creation of such pairs when we impose that momentum is to be conserved. But it does not seem that this would constitute a strong enough constraint under appropriate circumstances, because the pairs could be created without much momentum or through an input of momentum from the environment, as is the case for ordinary particle-antiparticle creation processes arising from the disintegration of a single boson. In section 1.11 I will examine the question of momentum and energy conservation more specifically, but for now it suffices to mention that when all contributions are taken into account it becomes clear that it is not the requirement of momentum conservation which prevents pair creation processes involving particles with opposite energies from occurring.

The fact that the kind of creation (or annihilation) processes which would require no energy input (or output) could be described as processes in which a particle reverses its direction of propagation in time while retaining the sign of its energy, may suggest another explanation for why such events would be forbidden. Indeed, we may ask why it is that when a particle changes its direction of propagation in time in the course of all those particle-antiparticle annihilation processes which do occur under the right conditions, the energy is invariably reversed relative to the new direction of propagation in time (so that it appears to be unchanged from the forward time perspective)? Why must it be imposed that a reversal of the direction of propagation in time be combined with such a reversal of energy which leaves the sign of action invariant, so that the energy of the annihilating pair needs to be compensated by the emission of photons carrying away the energy? Could it be that it is a requirement of continuity of physical properties along the world-lines of elementary particles that prevents a positive action particle from turning into a negative action particle? Such a change would in effect involve the transformation of a particle experiencing the gravitational interaction in a given way into a particle experiencing it in a different way, but perhaps that a particle cannot change the way it gravitationally interacts with the rest of the universe on a continuous world-line.

I must acknowledge that I once contemplated the possibility that action

sign changing reversals of the direction of propagation in time may be forbidden by a requirement of continuity of physical parameters along a particle's world-line. But I later came to understand that what such a requirement of continuity imposes is merely an absence of interruption of the flow of the fundamental time direction parameter, which can be satisfied even when the energy of a particle does not reverse upon a change of its direction of propagation in time. In section 2.10 I will explain what constraint a condition of continuity of the flow of time along an elementary particle wold-line would impose on the transformation of physical parameters and it will be clear that a reversal of the action is not forbidden by such a requirement. In any case, if the *charge* of a particle can change discontinuously (can reverse) from the forward time viewpoint when the particle reverses its direction of propagation in time in a continuous fashion (during a process perceived as an ordinary particle-antiparticle annihilation process), then there is no a priori reason why the *action* of a particle could not reverse in a similar manner when it reverses its direction of propagation in time, if the reversal in time also occurs in a continuous way, which would simply mean that the particle does not actually experience the usual reversal of its energy sign at the bifurcation point when it reverses its direction of propagation in time.

Actually, I believe that the simple fact that two opposite action particles of the same type must be considered to consist in the same particle which simply happens to be in a different energy state (or to propagate in a different direction of time) means that such particles should be allowed to transform into one another on a continuous world-line if their similarity is to ever be explainable in a causal way, but this is precisely what must occur only in rare circumstances. Must one then conclude that there exists an unexplainable decree simply banning negative action particles (carrying positive energy backward in time) from existing? This would again be the easy way out: there is a difficulty so let's just forget about the whole thing. But if we recognize that the existence of particles carrying positive energies backward in time is theoretically inevitable, then a satisfactory explanation for the absence of spontaneous matter creation is required.

Before dealing with the problem of matter creation I would like to address the related issue of the annihilation of pairs of opposite energy particles whose solution turns out to be much simpler than one could perhaps imagine. To understand what imposes a limit on the annihilation of pairs of opposite action particles we simply need to take into account the results obtained in the preceding section. Indeed, one may ask how it is supposed to occur

that a positive action particle with positive charge, say, could annihilate with a negative action particle with negative charge if positive and negative action particles are to be considered as equivalent to voids in opposite energy portions of the vacuum. How could the two particles ever come into contact with one another and annihilate when annihilation is to be considered a kind of interaction and there is absolutely no direct interaction of any kind between opposite action particles? Had I taken the lesson learned while solving the problem of the nature of repulsive gravitational interactions more seriously I would have understood much more readily that what limits the annihilation of particles with opposite energy signs is the absence of any direct interaction between such particles combined with the weakness of the indirect gravitational interaction they do experience. Indeed, in the absence of any direct interactions between them, two opposite action particles cannot even come into contact with one another and therefore would not be able to annihilate one another. Even if they were to find themselves near one another, two opposite action particles with opposite charges could not merge and combine their physical properties to perhaps produce a final state of no energy, because they do not even experience the presence of one another directly.

It is true though that opposite action particles would, according to the results I derived in the preceding section, be subject to some indirect gravitational interaction as a consequence of the equivalence between the presence of a particle with a given energy sign and an absence of energy of opposite sign from the vacuum. But given that the gravitational interaction between two elementary particles is negligible under most circumstances, it must be concluded that the probability of observing the annihilation of opposite action particles is very low unless the energies involved are extremely high (of the order of the Planck energy). Thus, given that under ordinary circumstances opposite action particles are only subjected to indirect interactions which are so weak that their effects become visible only when large amounts of matter are involved (in which case the energy exchanges between individual particles are still negligible), we have to conclude that no annihilation of opposite energy sign particles back to the vacuum would occur at any observationally significant rate, even if negative energy matter was present in our region of the universe with a density comparable to that of positive energy matter. It should be the case, however, that in situations of very high energy density, like those encountered in the very first instants of the Big Bang, processes involving the gravitational interaction of elementary particles with opposite

action signs would be likely to occur and could actually give rise to the annihilation of pairs of particles carrying positive energies in opposite directions of time.

Given those conclusions one may perhaps be tempted to argue that the problem of the creation of pairs of opposite action particles out of nothing is also one that arises merely when we fail to recognize that there are no direct interactions between the particles forming such a pair. This argument is not valid, however, because in the case of matter creation we do not need the energy to be present beforehand and there is no a priori reason why the opposite energies of the particles which would be created could not be arbitrarily large, therefore allowing the process to occur through the indirect gravitational interaction that is allowed to take place between the two particles involved. It is true that, somewhat paradoxically, pairs of opposite action particles with lower energies would be more difficult to create, because such pairs would be subject to weaker indirect gravitational interactions and therefore would be less likely to respond to local perturbations in the gravitational field. But the problem is precisely that there appears to be no limit to the amount of (positive and negative) energy which could be produced in the vacuum by such pair creation processes, so that it seems that particles with very high opposite energies should be produced continuously, even under normal conditions. What I have come to realize, however, is that despite the apparently inescapable nature of the problem of energy out of nothing, the observed absence of opposite action pair creation processes can be quite easily explained without even having to invoke any independent constraint applying on the creation processes themselves.

I think that what prevents processes of creation out of nothing from having undesirable consequences is in effect simply the fact that given that the probability for such processes to occur increases when the magnitude of the energies of the particles involved increases (due to the fact that the strength of the indirect gravitational interaction between opposite action particles rises when the magnitude of their energies rises) it follows that any particle that is produced in such a way has enough energy to immediately annihilate with a similarly produced opposite action particle present in the vacuum within the very short period of time characteristic of quantum gravitational phenomena. In other words, what prevents pairs of opposite action particles from being permanently created out of nothing under ordinary circumstances is the fact that two conditions with incompatible requirements must be met at the same time, because when the processes are favored from the viewpoint of creation

they involve particles with large (positive and negative) energies, while when they are favored from the viewpoint of duration they involve particles with much smaller energies. As a consequence, the creation of pairs of opposite action particles from nothing is very unlikely to have any observable consequences and we would be justified to expect that under normal conditions it is not possible for matter to be *permanently* created out of nothing, even when energy would be conserved in the process, because when matter particles are produced in such a way they usually annihilate back to the vacuum within a very short time through the same kind of processes.

It remains, however, that matter creation out of nothing is not totally impossible, because on a sufficiently small scale processes of opposite action pair creation would actually take place in the vacuum, even if they would usually be followed by the subsequent annihilation of the particles so produced in the context where annihilation to nothing is also likely to occur on the scale of distance and energy characteristic of quantum gravitational phenomena. But despite the impossibility for matter to be created out of nothing under normal circumstances, it appears necessary to assume that during the Big Bang processes involving pairs of opposite action particles would allow matter to be permanently created as a consequence of the rapid expansion of space. This is because when the expansion is very fast over a sufficiently long period of time, as it must have been in the first instants of the Big Bang (for reasons I will explain in section 3.5), two opposite action particles created as a pair can move away from one another rapidly enough that they may no longer be able to annihilate back to the vacuum (given that on the scale of quantum gravitational phenomena the distance between the particles would have become too large and their energies too low), which would mean that the creation process has become permanent.

In fact, if matter cannot be considered to simply exist but must be created along with space and time at the Big Bang, then the occurrence of processes of creation of pairs of opposite action particles out of the vacuum would become an absolute requirement. But even if it was assumed that matter already existed prior to the Big Bang (as may be allowed by some quantum theories of gravitation) it seems that the processes of creation out of nothing which are continuously occurring on a very short time scale would have to be allowed to become permanent under the conditions which existed in the very first moments of the universe's expansion given that such creation processes are required in order to reverse the consequences of the annihilation of opposite action particles which would necessarily be occurring in the moments

immediately preceding the formation of the singularity (if the initial matter distribution is sufficiently smooth, as I will explain in section 3.9). The conclusion that the existence of negative energy matter does not give rise to creation out of nothing under ordinary circumstances is certainly significant, but I believe that the conclusion that it is nevertheless possible for pairs of opposite action particles to be permanently created without energy input under the most extreme conditions is even more significant, particularly from a cosmological viewpoint.

It must be understood, however, that even when opposite action particles are involved, the processes of pair creation and annihilation would have to involve elementary particles with opposite directions of propagation in time, just as is the case for ordinary particle-antiparticle creation and annihilation processes. This is the true requirement of the condition of continuity of the flow of time which will be introduced in section 2.10 and which can therefore be seen not to forbid all creation and annihilation processes of the kind that would involve opposite action particle pairs, but merely those among such processes which would actually involve an interruption of the direction of the flow of time along a particle world-line, as when two forward in time propagating particles with opposite energies (and opposite actions) would meet and vanish. In this context it is important to understand that the backward in time propagating negative action particles which may be created along with forward in time propagating positive action particles must in effect be those with charges opposite (from the viewpoint of an observer measuring them in the forward direction of time) those of the positive action particles, if charge is to be conserved during any process of creation and annihilation of opposite action particles. This will later be explained to be allowed by the necessary invariance of the sign of charge (relative to its true direction of propagation in time) under both action sign preserving and action sign reversing discrete symmetry operations.

1.10   The problem of vacuum decay

There is an unavoidable question that arises whenever one proposes that negative energy states may be physically allowed. What is it in effect that prevents particles from falling into those 'lower' energy states? It has been argued that positive energy matter particles may not be able to do so because they would first have to surmount the limit imposed by the irreducible value

of their positive mass. But that would clearly not prevent particles already in a negative energy state from reaching even 'lower' energy states and given that I'm here working under the assumption that negative energy matter can exist in stable form this would appear to be a serious problem. Under such conditions it would seem that if even a small amount of matter was to ever find itself in one of the available negative energy states this would give rise to a catastrophic process of creation of negative matter energy and positive radiation energy, because the matter would radiate energy in going from the 'higher' energy states (with negative values nearer to zero) to the allowed 'lower' energy states (with larger negative values) without ever reaching a minimum energy in which it could settle down. Thus, as I mentioned before, it would seem that if negative energy matter can exist we could produce an infinite amount of work by simply harvesting the positive energy radiation produced when negative energy particles fall into lower negative energy states. But given that quantum field theory already allows for states of negative energy to occur in limited portions of space it would seem that we have a very serious problem, even in the current theoretical context, because if negative energy can be made to exist under such conditions (which have already been produced in the laboratory) it should immediately collapse to even lower negative energies and in the process produce an arbitrarily large amount of positive energy radiation, while of course no such phenomenon has ever been observed.

The insights gained while studying the problem of matter creation discussed in the preceding section, however, provide the elements needed to tackle this additional difficulty from a different angle. We may recall in effect that according to the preceding discussion an important consequence of the absence of any direct interaction between opposite action particles is that it is actually impossible, under ordinary circumstances, for a particle to annihilate with its opposite action antiparticle counterpart, which is another way to say that an already existing particle cannot reverse its direction of propagation in time without also reversing its energy sign (relative to its new direction of propagation in time), therefore describing an ordinary particle-antiparticle annihilation process. But another perhaps less obvious consequence of the absence of any direct interactions between opposite action particles is that a negative energy particle cannot emit a real (by opposition to virtual) positive energy interaction boson regardless of what energy changes the original particle goes through, because the positive energy boson could not even have been into contact with the negative energy

particle it is assumed to transform.

Therefore, a negative energy particle could not gain negative energy at the expense of the production of a compensating amount of positive radiation energy and the same limitation also implies that a positive energy particle couldn't absorb negative energy radiation and diminish its own positive energy in the process. This constraint must apply even if such processes could occur without violating conservation laws when the energy change of the matter particle involved would be compensated by the emission of an opposite amount of radiation energy. But this means that even the emission of *positive* energy radiation by a positive energy matter particle could not occur in such a way that the positive energy particle could turn into a negative energy particle, given that this would imply that there would have been a direct interaction between the now negative energy matter particle and the positive energy radiation it would have released, while according to my analysis this must be considered impossible.

Thus, the same constraint which allowed me to conclude that a particle cannot change its direction of propagation in time without reversing its energy sign also implies that it is impossible for a particle to reverse its energy without reversing its direction of propagation in time (in which case the particle would not continue to exist with opposite energy in the future). The existence of such a limitation suggests that no interaction vertex involving particles with mixed action signs needs to be taken into account in determining the transition probabilities of quantum processes. This is a valid conclusion even if the merger of certain opposite action particle world-lines may be allowed under conditions where the gravitational field is very strong, as I explained in the preceding section, because such annihilation processes would not occur through the emission of gravitational radiation (especially since they need not release any energy at all) but merely as a consequence of the interaction of the two particles involved with their respective same-energy-sign vacuums. A certain limitation against the possibility of transitions to negative energy states therefore actually exists, because a positive energy particle cannot 'fall' into a negative energy state by releasing positive energy radiation. The only reversal of energy which may occur on a continuous particle world-line would have to involve a reversal of the direction of propagation in time, in which case the energy of the particle would no longer be negative relative to the forward direction of time and we would merely observe a conventional antiparticle in a positive energy state annihilating with the 'original' particle.

The limitation imposed on vertexes that they cannot involve particles with mixed action signs would therefore actually prevent a particle that is already in a negative energy state from falling into even 'lower' energy states by releasing positive energy radiation, because such negative energy matter could never have been in contact with the positive energy radiation it is assumed to emit. In fact, this explanation works both ways, as it is also true that a particle in a negative energy state could not 'gain' energy and turn into a positive energy particle by releasing a compensating amount of negative energy radiation, because the bosons so released could not have been emitted by the now positive energy particle with which they can have no contact. What must be understood, again, is that while the requirement of energy conservation may not alone forbid transitions involving a reversal of the sign of energy, the fact that those transitions would involve the emission or the absorption of radiation with an energy sign opposite that of the original particle actually prevents them from occurring in the context where a negative energy particle (be it matter or radiation) can only interact with a positive energy particle through the very weak indirect gravitational interaction which exists by virtue of the fact that a negative energy particle can be described as a void in the positive energy portion of the vacuum.

Yet it must be remarked that the constraint described here would not prevent the vacuum itself from decaying by creating pairs of very high opposite energy particles, given that when the (positive and negative) energies are high enough, indirect gravitational interactions are allowed to occur between opposite energy particles. In the previous section I mentioned that this problem occurs only when we fail to take into account the fact that when the opposite energies of the particles produced are large enough for the processes of creation out of nothing to be likely to take place, it is also large enough for the particles so produced to immediately annihilate back to nothing with other particles of opposite energy sign present in the vacuum. But given that one of the particles which would be produced during such a process would actually have a negative energy, it may seem that an explanation is needed as to why it is exactly that the creation of this particle is not favored from a thermodynamic viewpoint, which could perhaps make the reverse process less likely to occur. In this particular sense it may therefore appear that a certain aspect of the problem of vacuum decay remains unsolved.

I believe that the situation we have here is analogous to that which was faced upon the introduction of the Rutherford atom model, which was initially rejected despite its apparent empirical inevitability, because it was

assumed that the electrons in orbit around the nucleus would lose energy in the form of electromagnetic radiation and end up collapsing into the nucleus, while no such catastrophe was observed. But just like the Rutherford model it appears that negative energy states are unavoidable and thus a solution to the problem of vacuum decay that does not simply amount to reject the physical nature of those states must be provided. Based on the results achieved in the preceding sections I would like to suggest that the difficulties described here arise again from the fact that we ignore the requirements imposed by the necessary relational definition of physical quantities. Indeed, what is happening is that we are attributing a direction to energy variations without referring to a physical aspect from our universe relative to which that direction could be compared. In other words, we use an absolutely defined direction on the energy scale which we arbitrarily define as 'lower' and we attribute distinctive physical properties to energy variations occurring along that absolutely defined direction, despite the fact that it actually has no objective significance. This traditional assumption seems to be justified by the observation that, for positive energy states at least, there does exist a singled out direction on the energy scale that is related to the natural tendency for matter to disintegrate and to reach thermal equilibrium. This direction can be associated with a well-defined physical aspect of our universe which is the direction of time in which entropy is growing. In the absence of such a relationship we would have no motive to assume the existence of a preferred direction on the positive energy scale that would not necessarily be opposite any such direction on the negative energy scale.

However, when I examined what the motives are exactly that allow us to consider the existence of this objectively defined 'lower' direction on the positive energy scale, arising in relation to the direction of time in which entropy grows, I realized that there is absolutely no reason to assume that this direction on the energy scale can be extended into negative energy territory without being subjected to a reversal like energy itself. The only assumption necessary to assert the validity of this conclusion is that the thermodynamic arrow of time points in the same direction from the viewpoint of both positive and negative energy observers, which certainly constitutes a plausible hypothesis, especially in the context of the explanation that will be proposed in chapter 3 for the origin of time asymmetry. Therefore, it seems that the objectively defined 'low' energy direction on the positive energy scale cannot be extended into negative energy territory, but would actually be effective toward smaller, *less negative* states (toward the zero energy ground state)

for negative energy matter.

Basically, what allows me to conclude that the low energy direction for negative energy matter is toward the zero energy, as is the case for positive energy matter, is that the singled out, objectively defined direction on the energy scale is simply that relative to which the energy tends to dissociate itself and to become less concentrated, so as to spread into a larger number of independent particles which thus necessarily have smaller (nearer to zero) energy as time goes. What explains this tendency is the fact that such a final configuration is associated with a larger number of microscopic degrees of freedom and a higher entropy (when gravitation can be neglected) and therefore is more likely to be reached in this direction of time in which entropy is actually allowed to grow. But, if the direction in time of entropy growth is the same for positive and negative energy systems, then the direction that would emerge as the low direction on the negative energy scale would have to be the opposite of that which constitutes the equivalent objectively or relationally defined low direction on the positive energy scale, because the spreading of energy into a larger number of particles with smaller *negative* energies, which is necessarily associated with a higher entropy, occurs in the direction on the energy scale opposite that in which smaller positive energies are reached. Thus, what we traditionally called 'low' energies, far below the zero point of vacuum energy, are in fact high energies for negative energy matter and what we called 'higher' energies, nearer to the zero point on the negative energy scale, are actually lower energies for negative energy matter. This is in perfect agreement with the previously discussed requirement to the effect that there should be a symmetry under exchange of positive and negative energy matter, so that the sign of energy can be defined as a relational property.

Such a conclusion is significant, because it allows one to deduce that it is not to be expected that matter should have a tendency (arising from a thermodynamic necessity) to decay into more negative energy states past the zero energy level. Negative energy matter must be expected to have the same tendency as positive energy matter to decay to energy states which from the perspective of an observer made of such matter would be lower energies and therefore to produce a larger number of particles with smaller negative energies and reach for the vacuum ground state in the future direction of time. If matter was found in a negative energy state it would not have a natural tendency to decay in a direction on the energy scale which is actually upward for a negative energy observer. It would be incorrect to assume that negative

energy particles have a tendency to decay by spontaneously gaining negative energy through absorption of negative energy radiation as time goes, because such configurations are not thermodynamically favored, but are actually less likely to occur for the same reason that positive energy matter particles are not likely to reach states where energy would become more concentrated into fewer particles as a result of the absorption of positive energy radiation. As a consequence, regardless of the energy level in which a positive energy particle is to be found at a given time, it can only release radiation until it reaches the energy contained in its rest mass and if it disintegrates and loses its mass it is not to be expected that it would continue to decay by gaining more negative energy through absorption of negative energy radiation. Thus, the vacuum itself should not have a tendency to decay by producing particles with arbitrarily large negative energies through processes of creation out of nothing that would become thermodynamically favored over the associated processes of annihilation to nothing.

The unavoidable character of the conclusion that there is no preference for 'lower', more negative energy states means that there should be no continuous decay to more concentrated negative energy states, especially in the context where there already exists a constraint on the release of positive radiation energy by matter entering a negative energy state. It would not be possible, therefore, to produce a large amount of work by making use of processes during which particles would gain larger negative energies either by releasing positive energy radiation or by spontaneously absorbing negative energy radiation, despite the assumption that matter is actually allowed to occupy those negative energy states. I should finally mention that the fact that we observe no catastrophic collapse to larger negative energies under the conditions where small negative energy densities are routinely produced in a limited way (as when a negative pressure is observed between two parallel mirrors in a vacuum) is a confirmation of the validity of the conclusions discussed in this section.

Thus, the outcome of the progress achieved in the last three sections is that it is possible to conceive of a fully consistent interpretation of negative energy states that would allow to at least preserve the validity of the current framework of quantum field theory. Indeed, it would appear that what we obtain are two more or less independent frameworks describing two more or less independently evolving categories of systems with opposite energies, which interfere with one another only under those special conditions where it is possible for an observer of one energy sign to indirectly deduce the

existence of opposite energy densities as they occur in the context where constraints are imposed which forbid the presence of certain states which would otherwise be present in that portion of the vacuum with the same sign of energy as that of the observer. This particularity allows the near perfect agreement between the predictions and the observations related to the small scale realm of quantum theory to naturally be maintained despite the fact that it is possible for matter to occupy the available negative energy states, which is also remarkable.

1.11   Energy and momentum conservation

I would now like to discuss the case of that most difficult of problems, which could have proved fatal to the alternative concept of negative energy developed here and which I have identified above as being that raised by the apparent possibility of a violation of the law of conservation of energy under conditions where interactions (even if merely of the indirect kind envisaged here) are allowed to occur between positive and negative energy matter. The nature of the issue can be illustrated through the use of a simple thought experiment. I briefly discussed in a previous section the problem that would arise in the case where a 'collision' would occur between a positive energy body and a negative energy body. I explained that such a collision would involve a loss or gain of positive energy by the positive energy body that would not be compensated, but instead be made worse by the associated gain or loss (respectively) of negative energy by the negative energy body. This is because instead of witnessing a loss of energy by one particle that would be gained by another, as when two particles with the same energy sign collide, we would here seem to have equal variations of energy, either both positive or both negative, depending on which particle accelerates and which decelerates as a result of the collision. For example, a negative action body could lose negative energy, while the positive action body it repels would gain positive energy, resulting in a net overall increase of energy twice as large as the individual changes. It would then seem that energy conservation is not possible under such circumstances.

The problem discussed here is also apparent when we consider the variations of momentum involved in such a process. Indeed, if action is to be assumed negative for a body propagating negative energy forward in time then it means that the sign of its momentum relative to its direction of prop-

agation in *space* must be negative, that is, momentum must be opposite the direction of the motion for a negative energy particle (because action has the dimension of an energy multiplied by a time or that of a momentum multiplied by a distance). In such a context it is easy to deduce that the variation of momentum occurring upon a collision between two opposite energy bodies would be twice as large as the absolute values of the changes in each particle's momentum rather than be zero as when two positive energy bodies collide. This is a problem that does not exist in the context of the traditional conception of negative energy matter according to which positive energy bodies attract negative energy bodies which repel them (if we assume that only gravitational forces exist between opposite energy bodies) and therefore the existence of such a difficulty could be used as an argument in favor of this traditional viewpoint despite the fact that it also raises other problems of its own, as I previously explained.

But given that we now understand that there are no direct interactions between opposite energy particles we have to recognize that the only way a collision between opposite energy bodies could occur would be through the indirect gravitational repulsion that would arise as a consequence of what are actually attractive gravitational forces attributable to a surrounding energy distribution, which are made to exist as a consequence of the equivalence between the presence of matter of one energy sign and an absence of energy of opposite sign in the vacuum. In this context it should in fact appear unlikely that there could occur violations of energy conservation arising from a collision between positive and negative energy bodies, if indeed there are no direct interactions between such objects. Mathematically at least, it certainly seems that a general relativistic theory of negative energy matter which would involve only gravitational interactions should not give rise to violations of the law of conservation of energy, given that energy conservation in such a context is actually a constraint concerning the exchange of energy between matter and the gravitational field.

Thus, if opposite energy bodies do interact only through the gravitational interaction, as I'm proposing, then it means that from the viewpoint of a general relativistic description of those interactions any variation in the energy of matter would in effect come from a variation in the energy of the gravitational field. The absence of any direct non-gravitational interaction between positive and negative energy bodies should indeed allow one to expect that it would be variations in the energy of the gravitational field that would balance the variations of energy occurring in the course of the interaction of

such opposite energy bodies. The problem I initially had, however, is that I was not able to figure out how this could come about in the more intuitive context of a Newtonian description of such interactions and I'm always suspicious of conclusions drawn solely on the basis of mathematical deductions, which often conceal totally inappropriate assumptions. So, where exactly does the positive energy go which is lost by a fast moving positive energy body colliding with a negative energy body initially at rest and where does the negative energy come from which is gained by the negative energy body that is accelerated during such a collision?

I was allowed to understand what is going on when a positive energy body interacts with a negative energy body only when I became aware of the possibility that the energy of matter and its gravitational field may be null for the universe as a whole. Indeed, as certain authors now recognize, it appears that when matter collapses to a spacetime singularity its gravitational potential energy becomes equal in magnitude (with opposite sign) to the energy of the matter itself. Thus, if the initial Big Bang state must be considered to consist of a spacetime singularity (which is required even in the presence of negative energy matter for reasons I will discuss in chapter 3), then it means that the gravitational potential energy of positive energy matter was initially the exact opposite of the energy of this matter. As space expanded this potential energy immediately began to decrease (toward the zero value) along with the positive kinetic energy of expansion, but it remains that under such circumstances there naturally occurs a compensation between the energy of matter and its gravitational potential energy (although it is actually the kinetic energy of expansion that must compensate the gravitational potential energy at all times, as I will explain in section 3.5). Even in the later stages of the evolution of a flat universe, when the gravitational potential energy of matter may appear to have become negligible, if a body gains kinetic energy and in the process acquires a large velocity relative to the average matter distribution, the potential energy attributable to the gravitational interaction of this body with the rest of the matter in the universe would rise to arbitrarily large negative values.

In such a context it is appropriate to assume that to any gain in the kinetic energy of a positive energy body there corresponds a similar but opposite gain in the energy associated with the gravitational interaction of this body with all the matter (with the same energy sign) in the universe. What I would like to suggest is that it would be plausible to assume that the required compensation for the kinetic energy gained or lost by a positive

energy body as a consequence of its indirect gravitational interaction with a negative energy body arises from a variation in the negative gravitational potential energy associated with the variation of positive vacuum energy that is equivalent to the related variation in the kinetic energy of this negative energy body.

Indeed, from my perspective, what is happening when a moving positive energy body indirectly communicates energy to a negative energy body is that while the positive energy body actually loses positive energy, the consequent gain in negative energy by the negative energy body is equivalent to a decrease in the amount of positive energy from the vacuum. But associated with this positive energy was a negative gravitational potential energy arising as a consequence of the interaction of this vacuum energy with the rest of matter and energy in the universe and if the above suggestion is right then this negative potential energy could be as large in magnitude as the positive vacuum energy which was present initially. Thus, the loss of positive energy by the positive energy body would be compensated by the loss of negative gravitational potential energy (which is a positive change) consequent to the reduction in positive vacuum energy which is equivalent for a positive energy observer to the energy increase (toward more negative values) experienced by the negative energy body. A similar reasoning also allows to conclude that the gain in negative energy experienced by the negative energy body is itself balanced by the gain in positive gravitational potential energy which follows from the increase in negative vacuum energy which is equivalent for a negative energy observer to the loss of positive energy experienced by positive energy matter (because a lesser amount of positive matter energy means a smaller void in the negative energy portion of the vacuum).

What must be understood here is that the reduction in positive vacuum energy which is equivalent to the gain in negative matter energy is actually a negative energy phenomenon and therefore does not have to be compensated by any change in positive matter energy or negative gravitational potential energy, which are positive energy phenomena (in the sense that they are associated with changes occurring in positive energy matter, or in the gravitational field between positive energy particles, or that between positive energy particles and the positive energy portion of the vacuum). Similarly the gain in negative vacuum energy which is equivalent to the loss of positive matter energy is to be considered a positive energy phenomenon that need not be compensated by a variation in negative matter energy or positive gravitational potential energy, which are actually negative energy phenomena (in

the sense that they involve changes occurring in negative energy matter, or in the gravitational field between negative energy particles, or that between negative energy particles and the negative energy portion of the vacuum). In any case if the above description is accurate then the energy that is lost or gained by a positive energy body as a result of its indirect gravitational interaction with a negative energy body could always be considered to be compensated by an opposite change in the gravitational potential energy associated with the variation of positive vacuum energy occurring as a consequence of the associated gain or loss of energy by the negative energy body.

I do recognize, of course, that under most circumstances the energy contained in the gravitational field associated with the interaction of a positive mass body with that portion of the surrounding vacuum with the same energy sign whose energy varies as a consequence of the equivalent variation of energy of the negative energy body with which the positive energy body interacts is much smaller than the energy change observed in the matter itself. But this does not mean that there is something wrong with the suggestion that the discussed variation in matter energy is compensated by some opposite change in gravitational potential energy, because the change in gravitational potential energy which I'm referring to here has to do with the interaction of this same portion of vacuum energy with the *entire* matter and energy content of the universe. Yet the fact that under all circumstances only as much energy as is present in a field of interaction can actually be exchanged between the particles interacting through that force field means that the energies exchanged during the process of indirect gravitational interaction between a positive and a negative energy body are relatively small and thus it is plausible that they could be compensated by a variation in some measure of gravitational potential energy associated with the changes involved. It must be clear, however, that we are not dealing here with the gravitational potential energy that could be associated with a repulsive force field mediating an interaction between the positive and negative energy bodies themselves, which in fact cannot exist as I explained before, but merely with *independent* measures of gravitational potential energy associated with the interactions occurring between those portions of the vacuum affected by the changes involved and the rest of matter and energy in the universe.

Thus, what must be understood is that following any interaction between a positive energy body and a negative energy body there actually occurs a variation in the total energy associated with positive and negative energy matter considered together, but this is only half of the equation, as to

any such change there must be a related compensating change in the gravitational potential energies associated with the equivalent variations in the positive and negative portions of vacuum energy. In the case of an interaction during which velocity is lost by a positive energy body and gained by a negative energy body, positive energy could actually be considered to flow from positive kinetic energy to positive gravitational potential energy, while negative energy flows from negative gravitational potential energy to negative kinetic energy. But it must be clear that this is only a reflection of the compensating opposite energy changes occurring in positive energy matter and its associated gravitational field on the one hand and in negative energy matter and its associated gravitational field on the other, because there is no actual exchange of energy between those two kinds of matter. It must also be mentioned that the variation in the momentum of matter which would be observed during such an indirect interaction is also compensated by the opposite variation in the momentum associated with the gravitational fields which occurs as a consequence of the changes in vacuum energy which are equivalent to the changes in the energy of matter. The fact that the gravitational interaction is very weak means that this energy flow between matter and gravitational fields is relatively small, but it nevertheless exists and it appears to be what allows energy to be conserved during such interaction processes.

1.12 Absolute inertial mass

One last objection which could be raised against the interpretation of negative energy states which I proposed has to do with the fact that from my viewpoint negative energy matter would offer the same resistance to acceleration as would positive energy matter. This would traditionally be described as being a consequence of the alternative assumption that inertial mass is positive even for negative energy matter otherwise characterized by a negative gravitational mass. Of course, as I already explained, the inertial mass must be considered to actually be reversed along with the gravitational mass from the viewpoint of a consistent description of the gravitational dynamics of negative energy matter. But in the context of the previously discussed improved conception of the phenomenon of inertia that emerged from my generalization of Newton's second law it was shown that acceleration would not occur in the direction opposite the applied force for a negative mass

body. In fact, once it is recognized that the equivalent gravitational field experienced by such an object must be opposite that experienced by a positive mass body, it is necessary to conclude that negative mass matter would actually experience the same resistance to acceleration as positive mass matter when submitted to the same forces, despite the reversal of its inertial mass. Thus, negative mass or negative energy matter would appear to violate the principle of equivalence as it is traditionally conceived.

Now, there could be situations where the gravitational mass in a volume of space would be relatively small or even zero despite the presence of a potentially large amount of matter in this volume, as when two opposite mass bodies are present all at once in the same location (which would be allowed in the absence of strong interactions between them). Such configurations would not be equivalent from an inertial viewpoint to the case of a system with nearly vanishing total mass, because the matter that is present would be more difficult to accelerate than if it actually had such a small mass. To better describe such vanishing energy configurations, which are clearly different from the vacuum, we may define a measure of inertial mass that would be related to the physically significant properties with which it is traditionally associated and that would correspond to the true amount of matter present under such circumstances, independently from the total amount of mass which may partially or totally cancel out. The *absolute inertial mass* obtained by adding the absolute values of the masses of all material bodies present in some volume of space (or by adding all masses as negative from the viewpoint of a negative mass observer) would constitute such a measure of the true amount of matter present.

It is clear that the acceleration of negative energy matter in a gravitational field attributable to a local matter inhomogeneity (such as the gravitational field which exists on the surface of the Earth) would not be that which is shared by all objects made of positive energy matter. Yet experiments provide very strong constraints on the degree of violation of the equivalence principle and to date there is in fact no evidence at all that any such violations have ever occurred when systems of various different compositions are utilized. However, I did say in a previous section that negative energy was as common as bound systems of particles such as atomic nuclei and molecules, due to the negative energy of their attractive force field. Why then do we never observe an altered level of resistance to gravitational acceleration? We may for example consider atomic nuclei formed of many protons and neutrons bound together by the strong nuclear interaction, with various measures of

negative energy of the force field associated with various configurations involving a variable number of component particles. It would then appear that the gravitational acceleration of such bound systems should be reduced by the negative value of the energy of the field while the inertial resistance would be proportionately larger, as the absolute inertial masses attributable to the component particles and the force field would not cancel out like the gravitational masses. If we measured the acceleration of a whole body composed of one such type of nucleus on the surface of the Earth and compared it with the acceleration of another body made of another kind of nucleus containing a lesser proportion of such negative energy, we may then perhaps expect to discern a difference. But it appears that this is precisely what the experiments discussed above rule out to a very good degree of precision. Shall we then once again abandon everything and conclude that negative energy, even though it is definitely present in bound systems, must be described in a non-relational manner (so that the sum of forces associated with inertial mass always cancel out like those associated with gravitational mass)?

It must be understood that in fact this conclusion would constitute a theoretical problem as grave as apparently is the empirical difficulty revealed by the absence of differences in the acceleration of various bound systems. Can we indeed ever hope to solve a problem by creating a 'new' one and assume that despite all indications to the contrary the latter difficulty is not real, simply because it only affects consistency on a more general level? This is not the path I chose to follow, because I realized that despite what is often suggested there is simply no reason to expect the kind of violations of the principle of equivalence which are described here, even if inertial forces do not cancel out when we consider two masses with opposite signs. What is wrong, I believe, with traditional assumptions is that when we are considering a bound system and its force field we assume that we have two masses with opposite signs, while what we really have is one single mass with one overall magnitude and one polarity, both from the viewpoint of inertia and from that of the response to local gravitational fields. Indeed, what motive would we have for considering that there could be independent contributions to the mass of a bound system (inertial or otherwise) when in fact the energy of the subsystems forming it (in particular the particles mediating the attractive force fields) could not be measured independently, given that they may be implicated in virtual processes which do not even have classically well-defined physical properties?

It is a fact that the particles mediating an interaction are virtual and

as such exist merely by virtue of quantum uncertainty, which allows them to carry energy, but only for a time that is short enough that this energy cannot be determined. The virtual particles involved in giving rise to interactions must then be considered unobservable, if only because to actually establish their presence in any one particular instance would require a time length greater than the duration of the exchange process. But under such circumstances how could we be talking about an *independent* contribution of those particles to the energy or the mass of the bound systems in which they materialize? I think that this would in effect be non-sense and that it must be recognized that any component of an entangled system whose physical properties cannot be directly and independently observed does not contribute independently to any of the properties associated with the mass of the entangled system as a whole, when those are actually measured. Failure to understand this decisive requirement would mean that we again allow one more inconsistency to obscure our conception of negative energy in a way that could only be made acceptable by rejecting one or another of the fundamental constraints identified above. In the present context this could not even be avoided by assuming that negative energy does not exist at all, because the issue is no longer merely about deciding if negative energy exists, but about determining its properties in a context where we must definitely accept that it is occurring.

There is no contradiction here, because there is definitely a negative contribution to the energy of bound systems, only this energy contribution cannot be independently measured in any specific case and this is the crucial distinction we must take into account when estimating the absolute inertial mass of such a system. Thus, the difference between the situation described above of the two superposed opposite mass objects with large absolute inertial masses and that of a composite system with absolute inertial mass smaller than that of its constituent particles is that in the former case we are actually dealing with two independent systems which may be interacting only negligibly with one another, while in the latter case we have a single entangled system which is physically different from the sum of its parts and to which must therefore be associated one single combined measure of mass, gravitational and inertial. In any case the fact that we do not observe violations of the principle of equivalence for bound systems whose observable total energy is positive confirms that this conclusion is appropriate.

1.13   A few other misconceptions

Before finishing this discussion concerning the potential problems facing a theory of negative energy matter I would like to provide arguments to the effect that a few other problems which are often associated with the possibility that there could exist gravitationally repulsive matter are actually of no concern, because they are significant only in the context of a traditional conception of negative energy and gravitational repulsion[6]. It is nevertheless important for me to discuss those issues, because I have come to realize that the perception of negative energy as being associated with all sorts of strange phenomena that defy common sense is responsible more than anything else for making the perfectly acceptable idea of negative energy matter look like a pseudo-scientific concept without any relevance to physical reality. I will thus try to make clear that what is wrong is not the hypothesis of matter in a negative energy state, but merely the current assumptions regarding what would be the properties of such matter.

One of the problems I would like to discuss arose as an outcome of the first attempts at finding an interpretation for the negative energy states which were predicted to occur by relativistic quantum theories. Indeed, when the existence of antimatter was experimentally confirmed it was suggested that this kind of matter may perhaps actually give rise to 'antigravity', in the sense that antimatter would experience repulsive gravitational forces in the presence of ordinary matter. But only theoretical arguments could be given to disprove this possibility when it was first suggested, because no experiment had yet been performed to demonstrate that antimatter would not fall upward in the gravitational field of the Earth. One of those arguments was based on the recognition that if antimatter was to repel or be repelled by ordinary matter this would allow perpetual motion machines to be build that would extract more energy from a process than was initially available. Indeed, under such circumstances it would take no energy to slowly raise

---

[6]It is not possible to provide a detailed review of all the papers which claim to offer a proof that gravitationally repulsive matter cannot exist in our universe, but I can assure the reader that even though I have carefully analyzed many of the so called 'theorems' against the existence of negative mass, or negative energy matter, I have never found any that does not contain one or another implicit or explicit assumption which would not apply to the kind of approach developed in this report and which invalidates them as theoretical arguments against the possibility of developing a consistent model based on the assumption that matter is allowed to occupy the available negative energy states.

a particle-antiparticle pair in the gravitational field of our planet (because there would be as much gravitational repulsion as attraction). But when this would be accomplished the pair could be made to annihilate and the positive energy of the photons so produced could fall back to a detector on the ground where they would be measured as carrying more energy than the pair initially had (this would be allowed in the context where the energy of the gravitationally repelled antiparticle is assumed to be positive relative to the forward direction of time) as a consequence of the frequency increase to which the positive energy photons would be submitted on their way down. It would then seem that energy can be freely produced if antimatter 'falls' up.

I think that this argument is perfectly valid, only it cannot be used to justify the rejection of anomalous gravitational interactions in general, but rather simply means that given that antimatter does not have negative energy (as observed in the forward direction of time) then it should not be expected to be submitted to anomalous gravitational forces. Now, could the same experiment be performed with negative energy (actually negative action) antimatter and then what would it mean for energy conservation? The answer to that question is to be found in the developments achieved by solving the problems discussed in the previous sections. First of all, it must be understood that given that there are no interactions between positive and negative energy matter other than the indirect repulsive gravitational interaction which I have already described, it seems that it would be much more difficult to raise a pair of opposite energy particles together in the gravitational field of a planet without doing work on at least one of them. Yet this may not constitute an insurmountable difficulty, because it is possible to imagine arrangements which would allow a negative energy body to achieve the task of raising a positive energy body in the gravitational field of a positive energy planet by making use of the indirect repulsive gravitational forces existing between the two bodies (which could also be composed of matter with opposite charges). But in fact the same limitation concerning the absence of any direct interaction between opposite energy particles would also imply that it is not possible to make such a pair to annihilate under normal circumstances, although again it is possible to imagine that the appropriate conditions to achieve this (a very high energy collision between two opposite energy particles) could perhaps be met when the appropriate technology would become available. However, other means would probably exist for harvesting the energy contained in each particle (or in each of the two

bodies) so that this limitation does not really constitute a decisive constraint that would allow to rule out the kind of processes discussed here.

The real difficulty for any incipient free energy harvesters would actually arise from the fact that in the context of a concept of gravitationally repulsive matter such as the one I proposed, even if a pair of opposite energy and opposite charge bodies could be raised together in the gravitational field of our planet without applying any external force on them, when the two bodies would annihilate they would release no energy at all. Indeed, if the objects have equal but opposite energies initially, they would not gain or lose any kinetic energy as a result of their ascension and this means that their respective final energies would still be equal in magnitude. As a consequence, even if their component particles could annihilate, no energy would be released, so that there would be no photons to fall back toward the surface of the planet with a net gain of energy. Of course we could arrange things so that the positive energy particles annihilate with other positive energy anti-particles already in place at the destination point, while the negative energy antiparticles would annihilate with negative energy particles already in place. But if the positive energy photons produced by the annihilation of the positive energy particles could actually gain positive energy while falling back to a detector on the ground, the negative energy photons for their part would lose negative energy while reaching the same detector and would therefore end up with less negative energy than they would have had if the negative energy matter had been submitted to annihilation before rising to a higher altitude. Thus, while positive radiation energy would be gained during such a process, negative radiation energy would be lost and this means that no useful energy can be produced in such a way.

In order to better understand the significance of the changes involved we can consider the variations occurring in the potential energy of the two bodies as they are raised in the gravitational field of the planet. From this more general perspective what would be observed in effect is that any potential energy that would be gained by one of the two bodies (the one that was actually lifted by the other) would necessarily be lost by the other body, thereby preventing any useful energy to be produced as a result of such a process. Indeed, while the positive energy body would gain positive potential energy (due to a loss of negative gravitational potential energy) the negative energy body would lose negative potential energy (due to an equivalent gain of positive gravitational potential energy). Now, this may seem to imply that a forbidden net increase of (positive) energy can be obtained despite the fact

that no work would have been done to take the system to its final state. Yet, as I have explained in a preceding section this variation is not significant, because any change in the energy of matter resulting from an interaction between positive and negative energy bodies is compensated by an opposite change in the energy of the gravitational fields associated with the equivalent variations in the positive and negative portions of vacuum energy.

What must be understood here is that even if there may occur changes in the potential energy of matter this would not mean that we have gained the ability to perform more work, as would be required to produce perpetual motion, because what the loss of negative potential energy by the negative energy body means is precisely that there was a loss of useful energy (energy that could be used to do work) for that object during the process by which it would have performed work to raise the positive energy body and increase the ability of this positive energy body to perform work. In other words, despite the net gain in potential energy for the pair as a whole, the ability to do work would not have increased, because the negative energy body, having been raised by the repulsive gravitational field it experiences, would now have a decreased potential to perform work (even though its kinetic energy would remain unchanged), which is precisely what its loss of *negative* potential energy implies, because indeed the object would have lost energy of the same sign as its own and therefore would actually end up with less energy available to perform work after the lifting process has occurred. The gain in useful energy by the positive energy body would actually have been provided by the negative energy body which would have lost its own useful energy and in fact, if the usual friction and other degradation of energy had been taken into consideration, it should be observed that the positive energy body would have gained less useful energy than the negative energy body would have lost, thereby precluding any perpetual motion from being achieved.

The fact that positive energy seems to have been created on the other hand is a simple consequence of the fact that the process discussed involves an indirect gravitational interaction between the two bodies and between the negative energy body and the positive energy planet during which the total energy of matter may indeed vary, as I remarked above, given that it is compensated by an opposite variation of the gravitational potential energy associated with the equivalent changes occurring in the energy of the vacuum. No additional difficulty is involved here and therefore it seems that the perpetual motion argument against gravitational repulsion cannot be considered significant other than as an argument against the possibility of an

anomalous gravitational interaction between ordinary matter and ordinary antimatter.

A more exotic and hypothetical phenomenon which according to certain accounts could have interesting practical applications, but which would raise serious problems from a theoretical viewpoint, given that it may provide the means of achieving faster than light space travel and therefore also time travel, is that of wormholes. It is often thought that wormholes would naturally occur in the presence of some types of black hole singularities and may allow remote regions of space to be directly connected in some way, so that traveling through such wormholes would enable to bypass the limitations associated with the passage of time experienced under normal circumstances when traveling over such long distances at slower than light velocity. It is not clear exactly what regions of space could be connected in such a way or if we are really talking about connecting regions of our own universe, but if we leave aside those uncertainties then it would seem that all that is required for unlocking the potential of faster than light space travel is the existence of traversable versions of such hypothetical shortcuts through space and time. What must be provided therefore is a means to maintain the 'throat' of a wormhole open for a long enough period of time that space travelers can safely traverse it despite the tendency for the matter configurations involved here to collapse under the effect of the gravitational attraction exerted by the singularity. The idea is that gravitationally repulsive negative energy matter (often called exotic matter) may allow to achieve that goal, given that it could be used to exert a gravitational repulsion that would compensate the attraction exerted by the spacetime singularity at the center of the black hole. But again, when we look at the details of such proposals, it becomes clear that the conditions necessary for achieving the desired results are incompatible with a consistent notion of negative energy matter. That may not be good news for science fiction lovers, but if I'm right negative energy matter could never be used to achieve such a goal.

To help identify what's wrong with current expectations I would suggest that we ask how it is exactly that negative energy matter could be brought not just inside some black hole, but toward the point of maximum density of positive energy matter (the singularity), despite the enormous gravitational repulsion that this positive energy matter would exert on the exotic matter? It should be clear that it is merely because we traditionally assume that negative energy matter would be attracted by a positive energy black hole and its singularity, even while it would repel it, that this appears to consti-

tute an achievable goal. But the truth is that any negative energy matter approaching a large concentration of positive energy matter such as an ordinary black hole would be submitted to repulsive forces as large as those maintaining positive energy matter trapped inside the same black hole. In this context the only way by which negative energy matter could find itself inside the event horizon of a positive energy black hole would be by having already been present inside the region destined to collapse into that positive mass black hole before it formed. But even if that was to happen there is no way that the negative energy matter could be made to remain near the black hole singularity where repulsive forces would be the largest. This situation is simply unstable and given that stability is precisely what is required for a traversable wormhole to exist, we must recognize that negative energy matter could not provide the necessary element for allowing spacetime singularities to be used for faster than light space travel and time travel. The possibility that the kind of phenomenon discussed here could actually have been used for achieving theoretically problematic, causality violating processes may seem far-fetched, but I think that it is nevertheless important to show that even under such extreme conditions there is no reason to expect that the hypothesis of the existence of negative energy matter could facilitate the occurrence of such self-contradictory phenomena.

The same argument I have used to rule out the possibility of engineering traversable wormholes can also be utilized to solve a more down-to-earth problem that is not often discussed, but which would contradict one of the most unavoidable constraint applying to the evolution of physical systems with a large number of microscopic degrees of freedom such as black holes. The problem is that negative energy matter, as it is traditionally conceived, could be used to reduce the mass of a black hole and therefore also the area of its event horizon. This could be achieved by simply throwing negative energy matter into a black hole, which would presumably absorb it given that negative energy matter is usually assumed to be gravitationally attracted by a positive energy black hole. This would be possible even if negative energy matter repels a positive mass black hole, because we could throw negative energy particles in small amounts and their gravitational fields would be too small to resist the much larger gravitational attraction of the black hole. But the surface area of a black hole has been shown to constitute a measure of the entropy of such an object, so that reducing the area of the black hole is similar to reducing its entropy. Again, however, if we reject the traditional conception of negative energy matter the problem does not exist, because a

negative energy particle cannot even get near a positive energy black hole without experiencing extreme gravitational repulsion, so that it certainly cannot be absorbed by the object, as would be necessary for reducing its mass and the area of its event horizon. If negative energy states are to be considered a true possibility then the fact that the traditional concept of negative energy matter would allow such violations of the second law of thermodynamics, while the alternative approach proposed in this report would forbid them, constitutes a strong indication to the effect that this latter description is more appropriate.

In fact, we are dealing with a much more general problem in this case, because from a traditional viewpoint it is actually assumed that when negative energy radiation would come into contact with positive energy matter (not necessarily a black hole) it could be used to withdraw positive thermal energy from this matter (as if it was providing negative heat), therefore again raising the possibility of allowing entropy to decrease as a consequence of the existence of negative energy matter. Of course given that from my viewpoint negative energy radiation cannot even come into contact with positive energy matter, the possibility raised here appears to be mostly irrelevant from a practical viewpoint. We may nevertheless examine the situation which would arise following an exchange of energy between positive and negative energy systems occurring as a consequence of the indirect repulsive gravitational forces they exert on one another.

The conclusion we must draw in such a case is that negative energy is not equivalent to negative heat for a positive energy system. Indeed, according to my conception of negative energy matter, kinetic energy is exchanged between opposite energy particles as if it was a positive definite quantity, which is allowed by the fact that the energy of matter is not conserved independently from certain contributions to gravitational potential energy associated with variations in the energy of the vacuum, as I explained before. But the fact that only the absolute value of the kinetic energy of matter is conserved means that thermal energy itself can only be exchanged as a positive definite quantity (or equivalently as a negative definite quantity from the viewpoint of negative energy observers) between opposite energy systems. Thus, when heat is provided by a negative energy system it can only raise the positive temperature of a positive energy system (as if positive thermal energy was provided) and the same is true for the heat provided by a positive energy system to a negative energy system which can only raise the negative temperature of the negative energy system (as if negative thermal energy was

provided by the positive energy system).

Thus, we have no reason to expect that even the indirect gravitational interactions between opposite energy systems could be used to transform useless forms of energy into more useful forms and in such a way reduce entropy. Negative energy cannot reduce the temperature of a positive energy system any more than positive energy could diminish (into positive territory) the thermal energy of a negative energy system, except under conditions where the magnitude of the temperature of one or another of two opposite energy systems is larger than that of the other system, in which case it is necessarily the system with the higher magnitude of temperature, regardless of its energy sign, that would lose positive or negative thermal energy and thereby raise the temperature of the other system by an amount proportional to that which is lost by the cooled system, as when only positive energy systems are involved. What must be understood is that transferring negative heat from a negative energy source to a positive energy system is not equivalent to removing positive heat from the same system. In fact, it turns out that adding heat from a negative energy system to a gas of positive energy matter can actually raise its temperature (unlike most people considering the possibility of the existence negative energy matter usually assume) instead of decreasing it. This is all a consequence of the fact that negative kinetic energy can be turned into positive kinetic energy and vice versa, even when energy is assumed to be conserved, as I mentioned above.

It appears, therefore, that the positive thermal energy of a gas of positive energy matter can actually be raised through contact with a gas of negative energy matter at a higher temperature (the temperature that would be measured by a negative energy observer) because thermal energy is a measure of the average kinetic energy of such a gas and this energy would become more evenly distributed (independently from energy sign) between the two gases if they could be put into contact through the indirect gravitational interaction. In this context it would appear that despite the fact that heat must be attributed a sign that depends on the sign of the energy that is gained or lost by a system, all that matters from a thermodynamic viewpoint for a positive energy system which interacts with a negative energy system is whether energy is actually gained or lost by the negative energy system and not whether the sign of this energy is positive or negative[7].

---

[7]For this reason a positive energy observer is allowed to consider temperature and heat as positive definite quantities under most situations when she is not dealing with the

Once again, the traditional expectation can be seen to arise from a misconception. You should take note, however, that I'm not just trying to debunk myths here. The opposite conclusion, that a low temperature gas made of positive energy matter would be cooled even further upon contact with negative energy matter or radiation, regardless of the magnitude of the temperature of this negative energy matter, and the above discussed assumption that a black hole's mass could be reduced through the absorption of negative energy matter, would constitute serious problems for a gravitational theory integrating the concept of negative energy matter. There are very strong motives behind my desire to demonstrate that the possibility of such entropy decreasing processes can be rejected and they are actually related to those which one might raise against the above discussed possibility of causality violating processes. I will explain what is the profound significance of the results discussed here in the multiple sections of chapter 3 that deal with the problem of time irreversibility.

1.14   An axiomatic formulation

Before I complete the process of integration of negative energy matter to classical gravitation theory I would like to provide formal statements of each of the significant rules I have derived in relation to this issue and which were discussed in the previous sections of the current chapter. Basically there are ten fundamentally decisive results which clarify the situation regarding the nature and the behavior of negative energy matter itself as well as the behavior of positive energy matter in the presence of negative energy matter. Those results actually provide the axioms or the rules on which a generalized classical theory of gravitation can be based. The axioms are legitimized by the fact that they have been shown to be necessary on the basis of both logical consistency and agreement with experimental facts and thus we may appropriately refer to them as principles. The first principle is the most fundamental and a recognition of its validity opens the way for a derivation of all the other results. The formal statement of this principle goes like this:

Principle 1: The distinction between a positive energy particle

---

thermodynamics of the gravitational field itself, as would become necessary when black holes are involved and the surface gravitational field is a measure of temperature. I will address the implications of attributing a negative temperature to negative energy matter configurations in the presence of strong gravitational fields in sections 2.12 and 3.7.

and a negative energy particle (propagating negative energy forward in time) can only be defined by referring to the difference or the identity of the energy sign of one particle in comparison with that of another, so that the sign of energy or mass has no absolute meaning.

From a gravitational viewpoint this principle is satisfied when positive energy particles are submitted to mutual gravitational attraction among themselves (as we observe), while negative energy particles (actually negative action particles) also attract one another gravitationally and positive and negative energy particles repel one another as a consequence of the indirect gravitational interaction which actually originates from an uncompensated gravitational attraction between matter of one energy sign and that portion of vacuum energy with the same energy sign. Compliance with this rule means that for a positive energy particle a negative energy particle should be physically equivalent to what a positive energy particle is for a negative energy particle. This property will be decisive for deriving the observer dependent generalized gravitational field equations that will be introduced later.

Another rule applies only in the classical Newtonian context where mass is a significant concept, but given that it allows to derive the rules which must also be obeyed in a general relativistic context it is necessary to mention it as a basic result. It simply amounts to recognize that:

> **Principle 2**: When mass is reversed from its conventional positive value both gravitational mass and inertial mass are reversed and together become negative.

This is actually equivalent to assume that there is indeed only one physical property to which we may refer to as being that of mass and that there cannot be any arbitrary distinction between gravitational and inertial mass.

While principles 1 and 2 are for the most part theoretically motivated the next principle is both theoretically and observationally motivated. Indeed, principle 3 arose as the unavoidable consequence of an analysis of the relationship between the attractive or repulsive nature of a field of interaction and the sign of the energy classically contained in this field, but it is also a necessary requirement of the fact that we do not observe any negative energy matter despite the fact that the existence of such matter appears to be allowed from a theoretical viewpoint. The third principle therefore is the following requirement:

> **Principle 3**: There are no direct interactions of any type (either gravitational, electromagnetic or any other), mediated by the exchange of bosons of interaction, between positive and negative action particles (respectively propagating positive and negative energies forward in time).

Compliance with this principle means that negative energy observers would also be prevented from directly observing positive energy matter.

Another important result was discussed at length in a previous section of this chapter where its validity was shown to be unavoidable despite the fact that it appears to contradict some assumptions which are usually considered to be irrefutable. This result simply states that:

> **Principle 4**: A void of limited size that develops in an otherwise uniform matter or energy distribution gives rises to uncompensated gravitational forces which are the opposite of those which would otherwise be produced by the matter or energy that is missing.

The effect it describes is the consequence of an alteration (caused by the presence of some local void) in the equilibrium of gravitational forces applying on any particle and due to its interaction with all the other particles in the universe (with which this particle actually interacts). The importance of this principle becomes clear when we consider its significance in the context where the uniform energy distribution is actually the distribution of vacuum energy and it is recognized that principle 5 below applies.

The following principle is probably the most decisive after principle 1 given that it is the result that allows the whole concept of negative energy matter to have a significance despite the validity of principle 3 and the absence of direct interactions between positive and negative energy particles. It states that:

> **Principle 5**: Locally, the presence of negative energy matter is equivalent to the absence of an equal amount of positive energy from the vacuum, while the presence of positive energy matter is equivalent to the absence of an equal amount of negative energy from the vacuum.

As explained in section 1.8 those equivalences constitute the particularity that allows opposite energy bodies to exert gravitational forces on one another despite the absence of direct interactions between them, simply because

according to principle 4 voids in a uniform positive energy distribution do have an indirect influence on positive energy matter despite the fact that those voids are actually equivalent to the presence of negative energy matter with which positive energy matter does not directly interact.

But even in the context where we assume the existence of a symmetry between positive and negative energy matter principle 5 would require that it is in fact only the inhomogeneities (either overdensities or underdensities) present in the negative energy matter distribution which can affect the gravitational dynamics of positive energy matter, while it is only the inhomogeneities present in the positive energy matter distribution which can affect negative energy matter. This is because, as previously discussed, the void in the positive energy vacuum that is equivalent to a totally homogeneous distribution of negative energy matter would leave no surrounding positive vacuum energy to produce an uncompensated gravitational attraction that would be equivalent (according to principle 4) to the gravitational repulsion otherwise attributable to the negative energy matter and the same is true concerning a homogeneous distribution of positive energy from the viewpoint of negative energy matter. An additional principle thus emerges that expresses this limitation applying on principle 5. It amounts to assume that:

> **Principle 6**: Only (positive and negative) density variations in an overall homogeneous cosmic scale distribution of negative energy matter can be assumed to exert gravitational forces on positive energy matter.

Of course a similar limitation would also apply which would actually express the absence of gravitational forces on negative energy matter from a totally smooth and uniform cosmic scale distribution of positive energy matter.

A further particularity could be derived from the already stated principles, but I will provide it as an additional specific rule because it may not be obvious that it applies in the context where principles 3 and 6 are assumed to constrain the interaction between positive and negative energy matter. This ordinance states that:

> **Principle 7**: Despite its energy sign and its assumed uniformity the negative energy portion of the vacuum does exert the traditionally expected gravitational influence it should have on positive energy matter.

As I previously explained this deduction (which would also apply to the positive energy portion of the vacuum from the viewpoint of negative energy matter) follows from the fact that the restriction that applies on the interaction of positive and negative energy matter does not prevent positive energy matter, when it is conceived as voids in the negative energy portion of the vacuum, from having an influence on that very portion of the vacuum in which the voids are present, just as voids in a matter distribution do exert an influence on this matter. Also, the fact that the energy of the vacuum may be expected to be uniformly distributed does not restrict the influence of the negative portion of it from influencing positive energy matter, simply because we are not dealing in this case with negative energy matter and the negative energy of the vacuum itself cannot be considered as being equivalent to a void in this very vacuum, so that whatever the extent of the distribution of negative energy involved it would still exert its influence on both positive and negative energy matter, unlike a uniform distribution of negative energy matter.

In a previous section I have explained that a consequence of principle 1 in the context where principle 2 (regarding the negativity of the inertial mass of a negative gravitational mass) is considered to apply is that the usual assumption that reversing all mass (gravitational and inertial) would allow to maintain agreement with the equivalence principle (as it is traditionally conceived) is wrong. Therefore, only an altered principle of equivalence between acceleration and a Newtonian gravitational field can remain valid. The additional condition applying on the equivalence principle would be the following:

> **Principle 8**: The equivalence of gravitation and acceleration does not apply merely locally, but merely for one single elementary particle (in a given location with a given sign of mass or energy) at once.

What remains true in this context is that the motion of bodies in a gravitational field does not depend on any physical properties of those bodies other than the sign of their mass or energy and this is what will allow the essence of the current theory of the gravitational field to be retained while accommodating a consistent concept of negative energy matter.

Another rule is observed in the context where negative energy matter is governed by principle 1 above and where the appropriate inertial behavior of this type of matter is assumed as a consequence of the validity of principle 2

and principle 6 (which actually imply that the inertial response of negative mass or negative energy bodies to a given force is the same as that of positive energy bodies, as I explained before). This rule would not be required if the traditional assumptions regarding the inertial response of negative energy or negative mass bodies were valid, but given that I have argued that those assumptions are problematic and cannot be justified then it seems that even traditionally we would have a problem if we were not taking the following experimentally motivated principle into account.

> **Principle 9**: When the negative contribution of a field of interaction to the energy of a bound physical system with overall positive energy cannot be independently and directly observed, only the diminished total energy of the bound system contributes to its (previously defined) absolute inertial mass.

Again this is also valid for bound physical systems with overall negative energy for which we may say that when the positive contribution of a field of interaction to the energy of the bound system cannot be independently and directly observed only the diminished (less negative) total energy of the bound system contributes to its absolute inertial mass. It must be remarked that the validity of this rule does not mean that the opposite contribution to the total energy of a bound system by the attractive field of interaction of its component particles cannot be well defined, only that if it cannot be isolated and independently measured then it also does not independently contribute to the inertial properties of the whole system.

One last constraint is observed to apply when negative energy states are allowed to be occupied (can be propagated forward in time). I have shown in a previous section that this rule can be considered to be theoretically motivated even though I initially deduced that it was necessary from purely phenomenological arguments. It is the following:

> **Principle 10**: In the absence of an appropriately strong local perturbation from the gravitational field a particle cannot reverse its direction of propagation in time without also reversing its energy and equivalently a particle cannot reverse its energy without also reversing its direction of propagation in time.

Here by 'negative energy' I mean negative energy relative to the true (even though arbitrarily defined) direction of propagation in time, as in the case of the positron as a negative energy electron propagating backward in time.

The ten principles enunciated above embody the essence of the insights I gained through an analysis of the problem of negative energy in light of the requirement of relational definition of the physical properties of mass and energy sign. They will now be used to help derive a generalized formulation of the gravitational field equations that will allow to describe the motion of particles with a given sign of energy in the gravitational field of an object with opposite mass or energy.

1.15   Generalized gravitational field equations

I previously indicated that equations would be scarce in this report. But the point has now been reached where it is absolutely necessary to provide some level of quantitative detail regarding the manner by which the concept of negative energy that was developed in the preceding sections is to be integrated into a classical theory of gravitation. The objective I'm seeking here though is not to provide a complete treatise on the subject, but merely to introduce the modified gravitational field equations which constitute the core mathematical structure of the generalized theory that emerges from the alternative set of axioms introduced in the preceding section. The essential requirement that must be imposed on a formulation of the gravitational field equations in the context where the principles enunciated in the preceding section are to govern the behavior of negative energy matter is that the gravitational field attributable to a given local source is not to be considered attractive or repulsive depending only on the sign of energy of the source. This can be satisfied by assuming that the gravitational field experienced by a negative energy particle and attributable to a given matter distribution is actually different from the one experienced by a positive energy particle. In such a context only the difference or the identity between the energy signs of two masses would be physically significant to determine the character of their gravitational interaction, so that any one mass could be considered to have positive energy while masses of opposite energy sign would then have to be the ones to which a negative energy is to be attributed. But the choice of which of two opposite energy bodies has positive energy is itself completely arbitrary.

Thus, an observer formed of matter with a given energy sign is free to attribute positive energy to particles with the same sign of energy, even though an observer formed of matter of opposite energy sign may attribute

a negative energy to the exact same matter. The only requirement is that the value of the gravitational field (which in a general relativistic theory is associated with the metric properties of space and time) always be adjusted as a consequence of the arbitrary choice which is made regarding the attribution of energy signs to various objects. There is, however, a natural choice for the attribution of energy signs by a given observer, which consists in assuming that matter with the same sign of energy as that of the observer itself is always to be considered positive by this type of observer. The viewpoint under which what we traditionally call positive energy matter actually has positive energy is therefore the natural viewpoint of what we traditionally consider to be a positive energy observer, while the viewpoint under which what we traditionally call positive energy matter actually has negative energy is the natural viewpoint of what we would traditionally consider to be a negative energy observer. When this convention is adopted we can write observer dependent gravitational field equations which replace the traditional equations. According to this alternative formulation the motion of matter with a given energy sign is determined by the gravitational field associated with observers having the same energy sign. The gravitational field therefore varies as a function of both the energy sign of the sources and the energy sign of the particles submitted to it, so that only the difference or the identity between the energy sign of the source and that of the matter submitted to the observer dependent gravitational field determines the repulsive or attractive nature of the interaction.

In a relativistic context the observer dependence of the gravitational field would imply that observers of opposite energy signs would actually experience space and time in a different way. But despite the awkwardness of this possibility from the perspective of our conventional perception of spatial relationships, from a mathematical viewpoint this requirement does not constitute an insurmountable difficulty. We merely have to assume two spaces, related to one another by the fact that the same unique set of events is taking place in both of them, but which may nevertheless have distinct metric properties, in the sense that the events which are taking place in the universe are separated by space and time intervals which are dependent on the energy sign of the observer. Indeed, as I mentioned before, the equations which will be proposed here merely constitute a generalization of the existing mathematical framework of relativity theory and we will therefore be in familiar territory. I'm in effect assuming that the reader already has a proper understanding of the current general relativistic theory of gravitation and of the

physical significance of the various mathematical objects which are relevant to the conventional formulation of this theory. Also, given that attempts at formulating a relativistic theory of gravitation that would allow for the existence of observer dependent gravitational fields were the subject of earlier publications by various authors and since it would be pointless to simply reproduce what has already been discussed elsewhere, I will leave to experts the task of introducing the general framework in which the developments I will propose are to be formulated and concentrate instead on describing the essential, distinctive mathematical features unique to the theory I'm proposing.

This choice is appropriate despite the fact that the approach I favor involves several distinctive aspects, because the most general features of the kind of framework involved are not dependent on the specific assumptions of the model considered. The reader may refer in particular to a relatively recent paper [24] in which were introduced meaningful developments essential to any theory according to which the gravitational field is assumed to be dependent on the nature of the matter experiencing it. But keep in mind that even the most suitable of the currently available mathematical frameworks still involves theoretical constructs and assumptions which I would consider inappropriate for the formulation of a fully consistent generalized classical theory of gravitation integrating the concept of negative energy matter and therefore only the general structure arising from those developments must be retained. I will here provide an interpretation of such bi-metric theories that is different from those which were tentatively proposed by the few authors that preceded me and this will have significant consequences which will be reflected in the fact that the final equations at which I have arrived are actually distinct from those which had been proposed until now.

In any case it must be mentioned that the gravitational field equations which appear in the above cited paper were not the first equations of that kind to have been developed. Gravitational field equations involving conjugate metrics had already been proposed that simply amounted to allow for negative contributions to the stress-energy tensor of matter[8], while implicitly conforming to the requirement of symmetry under an exchange of positive and negative energy signs. But even in the more recent publications no jus-

---

[8]I became aware of those developments mainly through the early writings of a Frenchman named Jean-Pierre Petit, but given that I have never read any official publication from him that contains the set of equations to be discussed here, then I will not attempt to provide specific references to his work on the subject.

tification has ever been provided for the assumptions on which are based the emerging theories and the only experimental consequences that were derived from those developments actually appeared to disagree with observations or were again unjustified on the basis of the hypotheses which were assumed to characterize the behavior of the gravitationally repulsive matter. In no case did the authors of those developments clearly recognized the exact nature of the anomalously gravitating matter they sought to describe, or attempted to explain how the various problems related to the existence of such matter could be solved. In fact, none of them succeeded in justifying the validity or the superiority of an approach to classical gravitation based on the requirement of exchange symmetry in comparison with the traditional viewpoint according to which gravitational attraction and repulsion are absolutely defined properties of matter.

Meaningful equations were nevertheless derived which happened to be compatible with the simplest of the conditions I have identified above as characterizing a consistent theory of negative energy matter. Those equations therefore constituted a step forward in deriving a quantitative model for the gravitational dynamics of negative energy matter, even if they failed to provide a totally appropriate framework and had to be assumed to apply only under particular circumstances, as they were clearly inappropriate to describe the early phases of cosmic evolution. In any case the equations which were initially proposed were of the following form:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -\frac{8\pi G}{c^4}(T_{\mu\nu} - T_{\mu\nu}^-) \tag{1.1}$$

$$R_{\mu\nu}^- - \frac{1}{2}g_{\mu\nu}R^- = -\frac{8\pi G}{c^4}(T_{\mu\nu}^- - T_{\mu\nu})$$

Here and in what follows G is Newton's constant, c is the speed of light in a vacuum, and the Greek indexes μ and ν run over the four general coordinate system labels (assuming a metric with diagonal elements $+1$, $+1$, $+1$, -1 in an inertial coordinate system). The usual notation is used for the curvature tensors $R_{\mu\nu}$ and R experienced by positive energy observers and for the stress-energy tensor $T_{\mu\nu}$ of what we conventionally consider to be positive energy matter. The curvature tensors experienced by negative energy observers are for their part denoted as $R_{\mu\nu}^-$ and R^-, while the stress-energy tensor of what we would conventionally consider to be negative energy matter is here denoted as $T_{\mu\nu}^-$. The first of those two equations can thus be used to determine the geodesics followed by positive energy particles, while the

second determines the geodesics followed by negative energy particles. Here all stress-energy tensors would have to be assumed to correspond with positive definite energy densities if it was not for the negative sign in front of the second stress-energy tensor on the right-hand side of each equation which allows a negative contribution to the total stress-energy tensor of matter that is dependent on the particular measure of the sign of energy associated with one or the other type of observer. The negative sign for stress-energies can thus be attributed alternatively to what we would usually consider to be negative energy matter and to what we usually consider to be positive energy matter.

This actually means that what appears to be negative energy matter to a conventional positive energy observer would really be positive energy matter for an observer we would normally consider to be a negative energy observer, while what appears to be positive energy to a positive energy observer would really be negative energy for an observer usually considered to be made of negative energy matter. Therefore, all energy signs must now be assumed to depend on the energy sign of the observer, which is itself merely a matter of convention. The viewpoint I previously identified as equivalent to a reversal of the sign of mass and according to which it is the gravitational field itself (represented here by the curvature tensors) which actually varies, while the sign of mass (replaced here by the sign of energy) of the observer which experiences that gravitational field is to be considered positive definite, is thus applied and this is certainly appropriate given that it gives rise to equations of the simplest form. It is because there are two different measures for the gravitational field, associated with the two different ways by which the positive and negative contributions to the total energy of matter can be attributed, that there are two equations for the gravitational field instead of the single one that is usually considered. Otherwise, however, those equations are fairly conventional and were certainly the most straightforward that one could derive for a bi-metric theory, as they were the closest to Einstein's own equation that one could propose.

The fact that, in the context of those equations, the sign of energy contributed by a given mass must now be assumed to depend on the sign of energy which we would normally attribute to the observer determining the associated gravitational field has important consequences. Indeed, if variations in the gravitational field (which is represented by the curvature tensors) are to compensate variations in the stress-energy of matter (as the general covariance of the equations require) then it means that the field attributed

to some matter can actually be either attractive or repulsive depending on the observer that measures the energy of this matter.

Four situations may therefore arise when we limit ourselves to merely permute the energy signs of a pair of interacting bodies. First, the source of the field could have what we traditionally consider to be positive energy and the field be attractive, because the particle submitted to it also has positive energy. Next, the source of the field could have what we traditionally consider to be negative energy and the field be repulsive, because again the particle submitted to it also has positive energy. Another possibility is that the source of the field could have what we would traditionally consider to be positive energy and the field nevertheless be repulsive, because we consider its effects on what we would traditionally consider to be a negative energy particle and from which viewpoint the source actually has negative energy. Finally, the source of the field could have what we traditionally consider to be negative energy and the field nevertheless be attractive, again because we consider its effects on what we would traditionally consider to be a negative energy particle and from which viewpoint the source actually has positive energy. This is certainly appropriate from the viewpoint of the principles identified in the preceding section. But given the insights I had already arrived at when I first learned about the mathematical developments which can be used to articulate those requirements, it appeared to me that what the available framework provided was at best an incomplete formulation of the gravitational field equations to be associated with a theory of negative energy matter.

To try to address those shortcomings I thus proposed (in a preprint [25] published in early 2006) the following equations which allowed to express the particularities of the indirect gravitational interaction of positive and negative energy mater that I had come to consider as unavoidable:

$$R_{\mu\nu}^{+} - \frac{1}{2}g_{\mu\nu}R^{+} = -\frac{8\pi G}{c^4}T_{\mu\nu}^{+} \tag{1.2}$$
$$R_{\mu\nu}^{-} - \frac{1}{2}g_{\mu\nu}R^{-} = -\frac{8\pi G}{c^4}T_{\mu\nu}^{-}$$

Here $R_{\mu\nu}^{+}$ and R^{+} are simply the curvature tensors experienced by positive energy observers while $R_{\mu\nu}^{-}$ and R^{-} are the curvature tensors experienced by negative energy observers. But the stress-energy tensors figuring in the equations I proposed are actually different from those entering the previously mentioned set of equations despite the similar notation I adopted here, because

the $T_{\mu\nu}^{+}$ tensor encompasses all contributions to the energy and momentum experienced by positive energy observers while the $T_{\mu\nu}^{-}$ tensor encompasses all contributions to the energy and momentum experienced by negative energy observers and I did assume contributions to those stress-energy tensors which were different from those which had previously been considered in the literature. Thus, when written in a more explicit form with all the components actually entering the stress-energy tensors on the right hand side, the equations I proposed are the following:

$$R_{\mu\nu}^{+} - \frac{1}{2}g_{\mu\nu}R^{+} = -\frac{8\pi G}{c^4}(T_{\mu\nu}^{+} + \grave{T}_{\mu\nu}^{-} - \acute{T}_{\mu\nu}^{-}) \qquad (1.3)$$

$$R_{\mu\nu}^{-} - \frac{1}{2}g_{\mu\nu}R^{-} = -\frac{8\pi G}{c^4}(T_{\mu\nu}^{-} + \grave{T}_{\mu\nu}^{+} - \acute{T}_{\mu\nu}^{+})$$

In this notation all energy-momentum tensors are assumed to be given in their positive definite form and now $T_{\mu\nu}^{+}$ is the stress-energy tensor of what is usually considered to be positive energy matter while $\grave{T}_{\mu\nu}^{-}$ is the stress-energy tensor associated with the measure of energy of negative energy matter below its average cosmic density (toward the zero energy level) and $\acute{T}_{\mu\nu}^{-}$ is the stress-energy tensor associated with the measure of energy of negative energy matter above its average cosmic density (toward more negative energies). Similarly, $T_{\mu\nu}^{-}$ is the stress-energy tensor of what we would usually consider to be negative energy matter while $\grave{T}_{\mu\nu}^{+}$ is the stress-energy tensor associated with the measure of energy of positive energy matter below its average cosmic density and $\acute{T}_{\mu\nu}^{+}$ is the stress-energy tensor associated with the measure of energy of positive energy matter above its average cosmic density.

This formulation of the generalized gravitational field equations allows me to take into account the fact that there are two distinct categories of contributions to the total energy density experienced by positive energy observers, one positive definite for all densities of positive energy matter and one that can be either positive or negative depending on the value of energy density of negative energy matter relative not to the zero energy ground state, but to the density of this negative energy matter averaged over the entire volume of the universe. Basically what that means is that the energy measures of the second category of contributions experienced by a positive energy observer are shifted from the traditional zero point of energy to a lower (more negative) energy level below which energies are negative and above which energies are positive up to a maximum value which is reached when no negative energy matter is present at all in the considered location. This redefinition of the

measures of energy associated with what we conventionally assume to be negative energy matter simply amounts to subtract the (time dependent) true, negative, average density of energy of this matter (add the absolute value of this density) from every measure of its energy density that contributes to determine the gravitational field experienced by what we conventionally assume to be positive energy matter, that is, the gravitational field observed by positive energy observers. I may add, however, that the required shift in the origin of energy measures for matter with an energy sign opposite that of the observer becomes significant only on the cosmological scale, because in the case of stars and planets it doesn't make much difference if we instead simply consider the true density of positive or negative action matter given that the typical densities which are then involved are much larger than the mean cosmic energy density, which can thus be neglected.

The refinement discussed here is justified (theoretically) by the fact that from the viewpoint of positive energy observers the description of negative energy matter as voids in the positive energy portion of the vacuum requires considering the contribution of negative energy matter as being merely relative to the average density of this matter distribution (and therefore to actually be positive in the presence of underdensities in the average cosmic distribution of negative energy matter) as a consequence of the absence of effects of a uniform negative energy matter distribution on positive energy matter which needs to be assumed for reasons I have explained in section 1.6. The equations I proposed also allowed to express the fact that a similar requirement exists for the contributions of positive energy matter to the total stress-energy tensor experienced by negative energy matter. But still I did not find the set of equations I had proposed completely satisfactory. I thought that the right solution should bring a simplification of the gravitational field equations, while visibly the equations I had derived were even less simple than the equations originally proposed by Einstein, despite the fact that in their compact form they were similar.

As I now understand, however, the equations I had proposed also fell short of meeting a certain mathematical requirement which I have come to appreciate as being essential to a consistent bi-metric theory of gravitation of the kind I sought to develop. This became clear when the paper [24] I mentioned above was published and new equations were proposed, apparently based in part on those I had developed, and which introduced a further refinement to bi-metric theories by not assuming that there is a unique predefined relationship between the metric properties associated with the measurements

of positive energy observers and those associated with the measurements of negative energy observers (even though for some reason the author of this paper preferred not to consider that the matter contributing a negative measure to the total stress-energy tensor experienced by positive energy matter actually constitutes negative energy matter). As a consequence of this revised assumption additional variables had to be considered that affected the contribution of negative energy matter to the total stress-energy tensor experienced by positive energy observers or the contribution of what we usually consider to be positive energy matter to the total stress-energy tensor experienced by negative energy observers. The equations proposed were the following, in which the additional factors are written in their explicit form, using my notation[9], and the quantities are now expressed in units where $c = 1$ and $G = 1/8\pi$:

$$R^+_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R^+ = -(T^+_{\mu\nu} - \sqrt{\frac{g^{-+}}{g^{++}}}a^{\nu}_{\nu}a^{\mu}_{\mu}T^-_{\nu\underline{\mu}}) \qquad (1.4)$$

$$R^-_{\nu\mu} - \frac{1}{2}g_{\nu\mu}R^- = -(T^-_{\nu\underline{\mu}} - \sqrt{\frac{g^{+-}}{g^{--}}}a^{\mu}_{\underline{\mu}}a^{\nu}_{\underline{\nu}}T^+_{\mu\nu})$$

The decisive additional factors are the determinants of what the author calls the pull-overs which are the maps $g^-_{\nu\mu}$ and $g^+_{\underline{\mu}\nu}$ (originally denoted $h_{\nu\mu}$ and $g_{\underline{\mu}\nu}$) which we may also write as $\boldsymbol{g}^{-+}$ and $\boldsymbol{g}^{+-}$ in tensor form. Those determinants are written here as $g^{-+} = \det(g^-_{\nu\mu})$ and $g^{+-} = \det(g^+_{\underline{\mu}\nu})$ while $g^{++} = \det(g^+_{\nu\mu})$

---

[9]From now on I will use a notation that allows to better represent the relative nature of the physical properties associated with spacetime and the gravitational field. In this notation tensors which refer to positive or negative stress-energies as determined from the viewpoint of positive energy observers will be given a plus or minus upper right index respectively. Tensors which refer to measures of spacetime curvature or metric properties as observed by positive energy observers will also be given an upper right plus index, while tensors which refer to the same kind of measures as observed by negative energy observers will be given an upper right minus index. Also when the distinct ordinary or underlined Greek letter indexes used in [24] are not explicitly present to show the nature of the tensor considered, I will simply add another plus or minus index to the right of that which already characterizes this tensor to define it as an object associated with physical properties as they are experienced by positive or negative energy observers respectively and associated with their own specific metric. For all such tensors, therefore, the first plus or minus index refers to the matter or gravitational field that is observed while the second plus or minus index refers to the matter that is observing. The underline which otherwise appears under some letter indexes can thus be considered as a shorthand for what should be additional plus or minus indexes over the letter indexes themselves.

is the determinant of the usual metric tensor related to properties of positive energy matter as observed by positive energy observers and $g^{--} = \det(g^-_{\underline{\mu\nu}})$ is the determinant of the metric tensor related to properties of negative energy matter as observed by negative energy observers (the map $\boldsymbol{a}$ is simply used as a means to transform the metric $\boldsymbol{g}^{++}$ into the $\boldsymbol{g}^{-+}$ pull-over or the metric $\boldsymbol{g}^{--}$ into the $\boldsymbol{g}^{+-}$ pull-over). It is clear therefore that the pull-over $\boldsymbol{g}^{-+}$ is the map which allows to describe the metric properties obeyed by negative energy matter as they are observed by positive energy observers, while the pull-over $\boldsymbol{g}^{+-}$ is the map which allows to describe the metric properties obeyed by positive energy matter as they are observed by negative energy observers (which justifies my notation). To better illustrate the relationships involved we may rewrite those equations as:

$$R^+_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R^+ = -(T^+_{\mu\nu} - \gamma^{-+}\sqrt{\frac{g^{--}}{g^{++}}}a^\nu_\nu a^\mu_{\underline{\mu}}T^-_{\underline{\nu\mu}}) \qquad (1.5)$$

$$R^-_{\nu\mu} - \frac{1}{2}g_{\nu\mu}R^- = -(T^-_{\underline{\nu\mu}} - \gamma^{+-}\sqrt{\frac{g^{++}}{g^{--}}}a^\mu_{\underline{\mu}}a^\nu_{\underline{\nu}}T^+_{\mu\nu})$$

where γ^{-+} is the absolute value of the determinant of the previously considered map of the metric properties of space experienced by negative energy matter as negative energy observers measure them to the metric properties of space experienced by negative energy matter as positive energy observers measure them and vice versa for γ^{+-}. We can then rewrite those equations in compact tensor form by making use of those *metric conversion factors* as:

$$\boldsymbol{G}^+ = -(\boldsymbol{T}^{++} - \gamma^{-+}\boldsymbol{T}^{-+}) \qquad (1.6)$$

$$\boldsymbol{G}^- = -(\boldsymbol{T}^{--} - \gamma^{+-}\boldsymbol{T}^{+-})$$

where $\boldsymbol{G}^+$ is the Einstein tensor $G^+_{\mu\nu} = R^+_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R^+$ related to positive energy observers, $\boldsymbol{G}^-$ is the similar Einstein tensor related to negative energy observers, $\boldsymbol{T}^{++}$ is the stress-energy tensor of positive energy matter as measured by positive energy observers, $\gamma^{-+}\boldsymbol{T}^{-+}$ is the stress-energy tensor of negative energy matter as measured by positive energy observers, $\boldsymbol{T}^{--}$ is the stress-energy tensor of negative energy matter as measured by negative energy observers and finally $\gamma^{+-}\boldsymbol{T}^{+-}$ is the stress-energy tensor of positive energy matter as measured by negative energy observers.

As is apparent, however, the proposed equations were still of the traditional kind, in the sense that they did not allow to take into account the

fact that negative energy matter is experienced as voids in the positive energy portion of the vacuum (and vice versa for positive energy matter from the viewpoint of negative energy observers). The complexity of those equations and their lack of symmetry under exchange of positive and negative energy states can be made more apparent by explicitly adding a term for the observed positive value of vacuum energy density:

$$\boldsymbol{G}^+ = -(\boldsymbol{T}^{++} + \boldsymbol{T}_\Lambda^{++} - \gamma^{-+}\boldsymbol{T}^{-+}) \qquad (1.7)$$
$$\boldsymbol{G}^- = -(\boldsymbol{T}^{--} - \boldsymbol{T}_\Lambda^{+-} - \gamma^{+-}\boldsymbol{T}^{+-})$$

In those equations $\boldsymbol{T}_\Lambda^{++} = -\Lambda\boldsymbol{g}^{++}$ would be the stress-energy tensor associated with the positive measure of energy density of vacuum fluctuations $\rho_\Lambda^{++} = \Lambda$ effected by a positive energy observer (with Λ as the positive cosmological constant measured by such an observer) while $\boldsymbol{T}_\Lambda^{+-} = -\Lambda\boldsymbol{g}^{--}$ would be the stress-energy tensor effected by a negative energy observer. The density of vacuum energy measured by a negative energy observer must be the opposite of that measured by a positive energy observer if the sign of energy is to remain an observer dependent physical property (which justifies the presence of a minus sign in front of the $\boldsymbol{T}_\Lambda^{+-}$ tensor that enters the gravitational field equations for negative energy observers). But given that we are indeed dealing with vacuum energy it would seem inappropriate to assign to this tensor the same metric conversion factor γ^{+-} as apply to measures of positive energy matter density performed by negative energy observers, even if the outcome of all positive and negative contributions to the energy of the vacuum is a positive energy, because in principle all such contributions exert a gravitational influence on both positive and negative energy observers on the cosmological scale. Anyhow, it is apparent that once all relevant contributions to the stress-energy tensors are considered, the symmetry of the original equations is lost, as their form becomes dependent on the actual sign of the average energy density of vacuum fluctuations. To me at least, it is obvious that those equations cannot be considered to embody a simplification of Einstein's theory that could be considered a substantial improvement over the original equations.

In order that such a formulation of bi-metric theory be allowed to at least meet the requirements I had already identified and which were not taken into account in this later proposal I would first suggest that we consider the limitations imposed on the interaction of positive and negative energy matter by the fact that the void of infinite extent in the positive energy portion of the

vacuum that is equivalent to the presence of a homogeneous distribution of negative energy matter has no gravitational effect on positive energy matter (and vice versa when we consider the similar void in the negative energy portion of the vacuum). In such a case we would simply have to replace the usual stress-energy tensors associated with the measures of energy of positive and negative energy matter made by observers of opposite energies with the following *irregular stress-energy tensors* which provide measures of the observed variations of energy density of positive and negative energy matter above and below their average cosmic densities:

$$\gamma^{-+}\check{\boldsymbol{T}}^{-+} = \gamma^{-+}(\boldsymbol{T}^{-+} - \bar{\boldsymbol{T}}^{-+}) \tag{1.8}$$
$$\gamma^{+-}\check{\boldsymbol{T}}^{+-} = \gamma^{+-}(\boldsymbol{T}^{+-} - \bar{\boldsymbol{T}}^{+-})$$

where $\gamma^{-+}\boldsymbol{T}^{-+}$ and $\gamma^{+-}\boldsymbol{T}^{+-}$ can be assumed to be the usual measures of stress-energy of negative and positive energy matter respectively (as experienced by observers of opposite energy signs) relative to the traditional zero level of energy and $\gamma^{-+}\bar{\boldsymbol{T}}^{-+}$ and $\gamma^{+-}\bar{\boldsymbol{T}}^{+-}$ are the measures of average stress-energy of negative and positive energy matter (as experienced by observers of opposite energy signs) observed on a cosmic scale. In such a context it appears that negative energy matter would now contribute negatively to the total measure of stress-energy experienced by a positive energy observer only when the magnitude of its local energy density (as measured by this positive energy observer) is larger than the magnitude of its average energy density (as measured by the same positive energy observer). Otherwise negative energy matter would actually contribute positively to the total measure of stress-energy experienced by a positive energy observer up to a maximum level fixed by the average density of negative energy matter (as measured by this positive energy observer). The same remark would apply for the contribution of what is usually considered to be positive energy matter to the total measure of stress-energy experienced by a negative energy observer, which would be opposite the energy contribution of negative energy matter only when the magnitude of the local density of positive energy matter (as measured by the negative energy observer) is larger than the magnitude of its average cosmic density.

It must be noted, however, that even though positive contributions to the energy density measured by positive energy observers may occur which would be attributable to the presence of underdensities in the negative energy matter distribution, we must nevertheless apply the conversion factor γ^{-+} to

such energy measures, because they still relate to measurements regarding the density of negative energy matter which are subject to the same mapping relationships as apply to other (truly negative) measures of energy related to negative energy matter made by a positive energy observer. Of course this is also true concerning below average measures of the energy density of what we would usually consider to be positive energy matter made by negative energy observers. Indeed, even when the second contribution to the energy density of matter is of the same sign as the energy of the matter experiencing the gravitational field it is still undetermined to the same extent as negative contributions, because what is unknown (due to the impossibility to directly compare measures of distances related to positive and negative energy observers) is the exact true density of negative energy matter (in comparison with that of positive energy matter) and this indefiniteness also affects the positive value of such contributions. Therefore, positive energy contributions from underdensities of negative energy matter are contained in the same tensor as negative energy contributions.

A more appropriate set of gravitational field equations would therefore take into account the shifted origin of the measures of stress-energy related to positive and negative energy matter as they are experienced by observers of opposite energy signs:

$$\boldsymbol{G}^{+} = -(\boldsymbol{T}^{++} + \boldsymbol{T}_{\Lambda}^{++} - \gamma^{-+}\boldsymbol{\check{T}}^{-+}) \tag{1.9}$$

$$\boldsymbol{G}^{-} = -(\boldsymbol{T}^{--} - \boldsymbol{T}_{\Lambda}^{+-} - \gamma^{+-}\boldsymbol{\check{T}}^{+-})$$

But clearly, for what regards simplicity, we appear to be no better off than with the previous set of equations. Something is still missing from those equations. At this point I suggest that we take a bold step forward and instead of trying to derive the gravitational field equations from a variational principle, as is usually done, we rather follow Einstein's way and simply guess what the final form of the equations should be that would generalize the set of equations (1.9) I have just proposed, which would otherwise constitute the most accurate description of the gravitational dynamics of positive and negative energy matter. As I have been able to understand, the crucial step in this process consists in reconsidering the meaning of the vacuum energy terms whose contributions I had long suspected were inappropriately attributed in the context of bi-metric theories. Indeed, I always thought that the cosmological term should arise from an asymmetry between some positive contribution and some negative contribution to the energy budget, while in

the current set of equations it occurs only as an additional term which must merely be given the appropriate relative measure depending on whether it is observed by a positive energy observer or a negative energy observer, which I do not find satisfactory.

It is only when I recognized the profound significance of my description of positive and negative energy matter as voids in their respective opposite energy vacuums that I was able to achieve the breakthrough that allowed me to guess what the appropriate generalized gravitational field equations are that allow the concept of negative energy matter to be integrated into a general relativistic framework in a way that actually simplifies Einstein's theory rather than further complicate things. What I realized, basically, is that if the results of my previously described analysis is right then all energy is vacuum energy, either present or missing. An additional insight was then necessary which consists in recognizing that the natural value of the positive and negative contributions to vacuum energy density is actually provided by the Planck scale. What must be understood is that when we remove energy from the vacuum we decrease its energy density from a maximum (positive or negative) value which is fluctuating quantum mechanically (upon measurement) in just the same measure as does the energy of matter itself. Therefore, if the presence of negative energy matter is to be considered as equivalent to the presence of a void in the positive energy portion of the vacuum, then locally we should observe a value of fluctuating vacuum energy density that would be decreased from its natural maximum value in just the same measure as that of the energy of the matter that is present. Given that the level of fluctuation of vacuum energy involved would be as large as the void considered is small it is possible to assume that there is an exact correspondence between the missing vacuum energy and the energy of the matter ordinarily expected to be present, which is known to be fluctuating (even if it is actually the measure of momentum that is involved) in proportion with the level of spatial confinement to which the matter is submitted. The natural energy level involved would thus correspond to that which is known to be associated with the highest level of fluctuation, which is actually the Planck energy[10].

---

[10]The validity of this assumption could be the subject of controversy, but given that the most advanced and least speculative theoretical developments toward a theory of quantum gravitation indicate that this is an appropriate and unavoidable constraint, I will nevertheless consider it to be universally valid. However, even if the existence of such a limit to the energy associated with quantum fluctuations was to be found irrelevant there is no a priori reason why the following results would have to be considered invalid.

Therefore, any missing vacuum energy attributable to the presence of matter with an energy sign opposite that of the portion of vacuum in which it arises could be considered as a local decrease over the maximum energy density determined by the Planck scale.

Let me thus introduce the generalized gravitational field equations which allow to fulfill all the requirements I have identified as essential aspects of a classical theory of gravitation that solves the problem of negative energies. The formula in all its beauty and simplicity is the following:

$$\boldsymbol{G}^{\pm} = -\boldsymbol{T}_v^{\pm} \tag{1.10}$$

where $\boldsymbol{G}^{\pm}$ is the Einstein tensor associated with the metric properties experienced by what we would usually consider to be positive and negative energy observers and $\boldsymbol{T}_v^{\pm}$ is the *vacuum stress-energy tensor* associated with the measures of vacuum energy effected by the same positive and negative energy observers. The similarity with the compact form of Einstein's own equation is very clear, but it is also somewhat misleading, as the right hand side of the equation proposed here is a much more general object than the stress-energy tensor of matter which appeared in the original theory. I will now define it with various levels of precision and generality. If we first consider the significance of the equation for a positive energy observer we would obtain the following equation:

$$\boldsymbol{G}^{+} = -(\boldsymbol{T}_v^{++} - \boldsymbol{T}_v^{-+}) \tag{1.11}$$

in which $\boldsymbol{G}^{+}$ is again the Einstein tensor associated with the gravitational field experienced by positive energy observers, but now the vacuum stress-energy tensor is decomposed into its positive and negative energy portions $\boldsymbol{T}_v^{++}$ and $\boldsymbol{T}_v^{-+}$ as they are measured or experienced by such positive energy observers. This is the most basic form of the proposed generalized gravitational field equations for a positive energy observer.

In accordance with what was explained above we would then obtain the next level of decomposition of the equations in which the two opposite energy portions of vacuum fluctuations (as they are experienced by positive energy observers) are given their explicit form:

$$-\boldsymbol{G}^{+} = (\boldsymbol{T}_P^{+} - \gamma^{-+}\boldsymbol{T}^{-+}) - (\gamma^{-+}\boldsymbol{T}_P^{-} - \boldsymbol{T}^{++}) \tag{1.12}$$

I believe that the situation we have here is similar to that which existed at the turn of the twentieth century concerning the hypothesis of the existence of atoms which was often rejected on the basis of an absence of direct observational evidence despite the fact that this assumption had actually become unavoidable theoretically.

where $\boldsymbol{T}_P^+$ and $\gamma^{-+}\boldsymbol{T}_P^-$ are the *natural vacuum-stress-energy tensors* associated with the maximum positive and negative contributions to the energy density of zero-point vacuum fluctuations which are directly experienced (other than through the gravitational interaction) by positive and negative energy matter respectively, but which both exert an observer dependent gravitational influence on positive energy matter. Here, the previously introduced metric conversion factors associated with the absence of fixed relationships between the metric properties of space experienced by negative energy matter and those experienced by positive energy matter are also applied to the maximum *negative* contribution to the energy density of vacuum fluctuations, given that it must be assumed (for reasons that will be explained in section 3.2) that this negative contribution is that which arises from the portion of vacuum fluctuations which are directly experienced by negative energy matter only. The γ^{-+} factors therefore occur only at the level of decomposition of equation (1.12) and not in equation (1.11), because they must in effect be attributed independently to the actual positive and negative energy contributions (of matter and vacuum) and the stress-energy tensors of matter provide energy contributions which are opposite those of the portion of vacuum in which they occur (given that matter of a given energy sign is to be conceived as voids in the opposite energy portion of the vacuum).

The preceding equation can then be rewritten under the following form when we take into account the previously introduced definition of the measure of stress-energy associated with negative energy matter as it would actually be experienced by positive energy observers, which are only affected by *variations* in the density of negative energy matter:

$$-\boldsymbol{G}^+ = \boldsymbol{T}^{++} - \gamma^{-+}\breve{\boldsymbol{T}}^{-+} + (\boldsymbol{T}_P^+ - \gamma^{-+}\boldsymbol{T}_P^-) \qquad (1.13)$$

which allows to identify the observed value of vacuum energy density associated with the cosmological constant observed by a positive energy observer as that which is provided by the following tensor:

$$\boldsymbol{T}_\Lambda^+ = \boldsymbol{T}_P^+ - \gamma^{-+}\boldsymbol{T}_P^- \qquad (1.14)$$

where the positive index attributed to this stress-energy tensor (associated with the energy of the vacuum in the absence of matter) now merely denotes the conventional energy sign of the observer experiencing it without referring to an actual energy sign of vacuum fluctuations themselves, which could in principle be either positive or negative (without affecting the form of the

equations) and which is determined solely by the conversion factor provided by the previously discussed map of the metric properties of space associated with negative energy matter as they are experienced by positive energy observers. Indeed, given the invariant nature of the maximum contributions to the density of vacuum energy associated with the Planck scale for an observer having the same energy sign as that of the contribution considered, the above equation means that the net value of vacuum energy density observed by positive energy observers arises as a consequence of a very small, but non-trivial difference in the metric properties of space associated with the motion of positive energy matter as experienced by positive energy observers and the metric properties of space associated with the motion of negative energy matter as experienced by negative energy observers. In any case it is now possible to write the generalized gravitational field equation associated with positive energy observers in its most explicit form as:

$$\boldsymbol{G}^+ = -(\boldsymbol{T}^{++} - \gamma^{-+}\boldsymbol{\check{T}}^{-+} + \boldsymbol{T}_\Lambda^+) \tag{1.15}$$

which confirms its formal equivalence with the first member of the previously proposed equation (1.9) at which I had arrived on the basis of considerations of a physical nature. It may be added that if we are considering this equation in a cosmological context then the $\gamma^{-+}\boldsymbol{\check{T}}^{-+}$ tensor would presumably reduce to zero on average (as the overdensities of negative energy matter would cancel out the underdensities present in the same matter distribution) so that the relevant equations for positive energy observers would now be of the traditional form:

$$\boldsymbol{G}^+ = -(\boldsymbol{T}^{++} + \boldsymbol{T}_\Lambda^+) \tag{1.16}$$

as is known to be appropriate given the success of current cosmological models for predicting the relevant features of our universe's history.

We may then also write the following set of equations which would provide the various levels of decomposition of the general equation (1.10) that apply from the viewpoint of negative energy observers:

$$
\begin{aligned}
\boldsymbol{G}^- &= -(\boldsymbol{T}_v^{--} - \boldsymbol{T}_v^{+-}) \\
-\boldsymbol{G}^- &= (\boldsymbol{T}_P^- - \gamma^{+-}\boldsymbol{T}^{+-}) - (\gamma^{+-}\boldsymbol{T}_P^+ - \boldsymbol{T}^{--}) \\
\boldsymbol{G}^- &= -(\boldsymbol{T}^{--} - \gamma^{+-}\boldsymbol{\check{T}}^{+-} + \boldsymbol{T}_\Lambda^-)
\end{aligned} \tag{1.17}
$$

where $\boldsymbol{T}_\Lambda^- = \boldsymbol{T}_P^- - \gamma^{+-}\boldsymbol{T}_P^+$ would provide the (positive or negative) value of vacuum energy density observed by such a negative energy observer. The

last equation, as well the other two, are now manifestly symmetric with the corresponding equations associated with positive energy observers, as I have argued should be required. But the most remarkable feature of those equations (and the related equations for the gravitational field experienced by a positive energy observer) is that they are actually obtained from a very simple expression (the first of the three equations) that determines the gravitational field merely as a function of the relatively defined measures of positive and negative vacuum energy and which alone allows to embody the essence of the emerging framework.

It must be noted that both the value of total vacuum energy density (associated with an absence of matter) that is measured by positive energy observers and that which is measured by negative energy observers (which are here given by the vacuum energy terms $\boldsymbol{T}_\Lambda^+$ and $\boldsymbol{T}_\Lambda^-$ respectively) could in principle vary with position (and incidentally also with time) given that they involve the variable metric conversion factors γ^{-+} and γ^{+-} respectively. Thus, the measure of vacuum energy density associated with the cosmological constant and applying on the global scale would actually be an average quantity and there is no a priori reason why it could not give rise to local effects that would deviate from those associated with the cosmic scale. In sections 3.2 and 3.3 I will explain how one must interpret the variable nature of the cosmological term and why it is still appropriate to consider that the density of vacuum energy does not vary with position in the absence of local inhomogeneities in the positive and negative energy matter distributions. Anyhow, given that we know that on the cosmic scale at least $\boldsymbol{T}_\Lambda^+ = \boldsymbol{T}_P^+ - \gamma^{-+}\boldsymbol{T}_P^-$ is very small compared with the natural energy scale provided by the Planck energy, then it is possible to conclude that the correction provided by the γ^{-+} conversion factor is itself actually very small on such a scale. This observation, therefore, indicates that there is a near perfect level of symmetry between the metric properties of space experienced by positive energy observers and those experienced by negative energy observers at the present epoch.

The quantitative aspects of the proposed integration of negative energy states to classical gravitation theory having being properly introduced, it is now possible to look back and examine whether the equations obtained can actually provide the structure of an alternative model which would conform to all of the principles enunciated in the preceding section. As I previously remarked the basic structure of the proposed bi-metric theory was adopted

precisely because it allows the kind of arbitrariness of the attribution of the sign of energy that is required for this physical property to be defined in a relational manner. But the ultimate confirmation that the proposed framework is compatible with the fundamental requirement expressed by principle 1 is the fact that even in the presence of a non-vanishing value for the cosmological constant, the set of equations (1.17) describing the motion of negative energy matter is now symmetric with the corresponding set of equations describing the motion of positive energy matter. Furthermore, the requirement set by principle 2 that inertial mass be reversed along with gravitational mass is also fulfilled by the proposed gravitational field equations given that my analysis of the physical property of inertia has shown that imposing such a condition should give rise to gravitational attraction between masses of the same sign (whatever this sign is assumed to be) and to gravitational repulsion between masses of opposite signs and this is precisely what we obtain with the proposed equations, even if the sign of energy that replaces the sign of mass is here arbitrary and the gravitational field is a variable property dependent on the nature of the matter submitted to it.

On the other hand, the validity of principle 3 and the absence of direct interaction between positive and negative energy matter particles may seem to be threatened by the fact that the stress-energy tensor associated with negative energy matter contributes to determine the gravitational field experienced by positive energy matter. But again, in the context of the more refined set of equations I have proposed, it is explicit that the negative contribution that enters the total measure of the stress-energy of matter that determines a gravitational field and which we associate with the presence of negative energy matter is actually a measure of the amount of stress-energy missing from the positive portion of vacuum energy. The effect on positive energy matter which must be taken into account in the presence of negative energy matter cannot therefore be attributable to an interaction with negative energy matter (whose presence is not directly felt by a positive energy observer), but must necessarily come from an interaction between positive energy matter and the surrounding positive energy vacuum. The equations thus naturally require that there are no direct interactions between particles with opposite energy signs.

The new equations are also the perfect embodiment of the requirements set by principles 4 and 5, because they allow the voids in the positive energy portion of the vacuum to actually provide a negative contribution to the total stress-energy tensor of matter and in a general relativistic context

a negative contribution to the stress-energy of matter must be matched by a contribution to the gravitational field that is opposite to that produced by positive stress-energy, so that if positive energy produces an attractive gravitational field from the viewpoint of positive energy matter, negative energy must produce a repulsive gravitational field from the same viewpoint. The presence of voids in an otherwise uniform distribution of positive vacuum energy should therefore give rise to uncompensated gravitational forces opposite those attributable to the presence of an equivalent amount of positive energy matter and by analogy the same should also be true for voids in a uniform positive energy matter distribution.

We can now understand why it would be inappropriate to assume, as some authors do, that the energy of the gravitationally repulsive matter whose behavior is described by conventional bi-metric theories is positive even for an observer that measures a negative contribution from it to the total stress-energy of matter. Indeed, according to the above proposed equations such matter would produce a gravitational field that would itself have an energy content (to the extent that a definite energy could actually be associated with the gravitational field) opposite that of the gravitational field which is produced by particles contributing positively to the total stress-energy of matter. But this means that if matter was assumed to always have positive energy, then when energy is exchanged between the two types of matter the variation of total gravitational energy (which would occur because opposite variations of *opposite* gravitational energies are involved) would not be compensated by a variation of the energy of matter (which would involve opposite variations of *positive* energies). Therefore, in the case of our two colliding bodies exerting a gravitational repulsion on one another it would be impossible for the variation of energy of the decelerating body to be compensated by a variation of energy of the gravitational field attributable to the changes occurring in the related portion of vacuum energy which would be equivalent to the energy changes occurring as a consequence of the acceleration of the second body, despite the fact that this must be considered necessary if energy is to be conserved, as I previously explained.

Those problems can be avoided, however, when real negative energy states are allowed for matter, because in a general relativistic context the variations in the gravitational field can actually balance the changes occurring in the stress-energy of the two interacting matter components and given that the gravitational interaction is responsible for all energy exchange between opposite energy bodies, then no energy variations remain uncompensated. I

think that this is a clear indication that the tentative solution to the problem of vacuum decay (the collapse of matter to ever more negative energy states) through the contradictory proposal of a gravitationally repulsive matter that would have positive energy (from all viewpoints) is misguided and ineffective. Thus, if an observer is allowed to attribute a positive energy to matter of his own kind, regardless of which matter he is made of, it should be clear that once this choice is made the energy sign of the matter which from the viewpoint of this same observer provides a negative contribution to the stress-energy tensor of matter must be assumed negative. In any case I must mention again that from a cosmological viewpoint the growth of negative energy matter overdensities occurring in an initially homogeneous distribution of such matter will always be compensated by an opposite growth of underdensities in the surrounding environment. But given that from my viewpoint those two kinds of inhomogeneities provide opposite contributions to the total stress-energy tensor of matter experienced by a positive energy observer, then it follows that there is an additional constraint regarding the conservation of energy contributed by negative energy matter and this is a further confirmation of the viability of the proposed equations.

Returning to the criteria imposed by the principles enunciated in the preceding section, we can readily assess that the condition set by principle 6 (according to which only density variations over and below the average cosmic density of negative energy matter have an effect on positive energy matter) is also reflected in the equations proposed above. Indeed, the modified measure of negative stress-energy provided by the irregular stress-energy tensor $\gamma^{-+}\breve{\boldsymbol{T}}^{-+}$ which naturally enters the gravitational field equation associated with a positive energy observer actually allows to fulfill the requirement set by principle 6 given that it provides a measure of stress-energy from which is subtracted the average stress-energy of negative energy matter. This compliance of the proposed gravitational field equations may perhaps appear to be of secondary concern given the negligibility of the average density of positive energy matter (and presumably also of negative energy matter) in comparison with the densities encountered under most circumstances when we are dealing with astronomical objects of interest like stars or even galaxies. But, if it was not for the modified measure of negative stress-energy provided by the second term of equation (1.15), or the corresponding term from equation (1.17), serious problems would occur.

In section 1.5 (in which was elaborated the alternative concept of negative

mass on which is based the mathematical framework developed here) I mentioned in effect that if a body with a given mass sign was to interact with all matter of both positive and negative mass that is present on the cosmological scale then the classical phenomenon of inertia itself could not even exist (because the inertial forces resulting from acceleration relative to positive and negative mass matter would cancel out). However, a Newtonian model is all about inertia, so that if inertial forces were made impossible by the presence of negative energy matter, then reduction of the relativistic equations to a Newtonian gravitation theory with gravitationally repulsive, negative mass densities would actually be impossible, even as an approximation. I believe that ignorance of the requirement to impose a suitable, modified measure of negative stress-energy for the generalized gravitational field equations is in fact the ultimate source of the difficulties which according to certain authors are encountered in trying to obtain an appropriate Newtonian limit from traditional bi-metric theories. This is in addition to the fact that, without the appropriate measure of negative stress-energy, complex hypotheses (of the kind which are often found in the literature) would have to be introduced concerning the variation in time of the ratio of the average cosmic densities of positive and negative energy matter in order to try to maintain the agreement of the proposed models with astronomical observations regarding the rate of expansion of ordinary positive energy matter, which is already predicted with good accuracy by traditional cosmological models when no negative energy matter is assumed to be present.

Finally, the fact that two maximum contributions of opposite signs to the energy density of the vacuum are now explicitly present in the most general form of each of the gravitational field equations means that both positive and negative contributions to the energy of the vacuum itself (ignoring voids) are allowed to contribute to the gravitational field experienced by positive or negative energy matter on the cosmological scale, as required by principle 7. From this alternative viewpoint what allows one to appropriately ignore most of the effects that the vacuum would have on the gravitational field experienced by positive or negative energy matter is merely the fact that those opposite energy contributions nearly cancel each other out at the present epoch. I may also mention that the condition set by principle 8 (that the equivalence principle be valid not merely locally, but really for one unique particle with a given energy sign) is implicitly contained in the structure of the equations at the most basic level, because they describe gravitational fields which are dependent not merely on the location, but also on the sign

of energy of the particles submitted to them. On the other hand, principles 9 and 10, which identify requirements that have to do with the properties of matter particles (namely the absence of independent energy contributions for bound systems and the impossibility under ordinary circumstances of a reversal of action on a continuous particle world-line), are not explicitly contained in the gravitational field equations proposed here, but if we assume the validity of those equations then experimental facts make those constraints unavoidable.

1.16   Summary

To conclude this chapter, I would like to provide a summary of all the results which were obtained concerning the problem of negative energy in the context of the improved understanding of the issue of time directionality which underlies those developments. The reader who may want to skip this section can do so without missing any essential development necessary to understand other portions of the present report. The decisive results are the following.

1. There is no valid observational argument against the existence of negative energy matter and what is required by the facts is merely that negative energy matter does not interfere under most circumstances with processes involving positive energy matter.

2. The introduction of antiparticles does not constitute a complete and acceptable solution to the problem of negative energy states.

3. There exists a fundamental degree of freedom associated with the direction of propagation in time of elementary particles.

4. The sign of energy is purely conventional given that it cannot be defined independently from the direction of propagation in time of the particle carrying this energy which is itself a matter of coordinative definition.

5. The only significant measure of energy sign from a gravitational viewpoint is that provided by the sign of action obtained by multiplying the sign of energy by the sign of time intervals.

6. The sign of action is also a matter of convention dependent on the choices made regarding what should be the sign of energy of those particles which are considered to propagate forward in time.

7. The sign of action cannot be asserted other than as a relative property between different particles sharing the same convention regarding what shall be the direction of propagation in time of a given particle with an arbitrarily chosen sign of energy relative to this direction of time.

8. As a consequence of a certain condition of continuity of the flow of time along an elementary particle world-line, a particle with a given conventionally defined sign of charge relative to a given direction of propagation in time cannot be allowed to also exist as a particle carrying an opposite charge in the opposite direction of time.

9. Any anomalous response of a conventionally defined negative action particle to the gravitational field of a conventionally defined positive action body must be shared by a positive action particle in the gravitational field of a negative action body.

10. As a matter of consistency a negative action or negative mass body must be assumed to have both negative gravitational mass and negative inertial mass, not because of some perceived requirement from the equivalence principle, but because mass as one single physical attribute cannot be assigned mutually exclusive or contradictory values.

11. Contrarily to what is usually assumed the hypothesis that inertial mass reverses along with gravitational mass does not give rise to absolutely defined attractive or repulsive gravitational fields.

12. It is the incorrect assessment of the response of a negative mass body to any applied force, made on the basis of current assumptions regarding the effect of a reversal of inertial mass, that is responsible for allowing an absolute character of attractiveness or repulsiveness to be associated with a given sign of mass.

13. It is inappropriate to assume that inertial mass remains positive for a negative mass body not only because this assumption would not give rise to the kind of ordinary response to forces that is usually assumed of such a mass, but also because even if the response was appropriate a body with such properties would irreconcilably violate the equivalence principle as a consequence of the fact that the same inertial mass would respond differently to a given gravitational field depending on the sign of the associated gravitational mass.

14. The direction of the equivalent gravitational field experienced by an accelerating body must be considered to be dependent on the sign of mass of the body and therefore to be equal rather than opposite the acceleration for a negative mass body, so that the inertial force on such a negative mass body is left invariant despite the fact that its inertial mass must be assumed to be negative.

15. A generalized formulation of Newton's second law involving a dynamic equilibrium between applied forces and the inertial force associated with the equivalent gravitational field, instead of an equilibrium between forces and acceleration, allows to predict that $\boldsymbol{F} = -m\boldsymbol{a}$ when the mass m is negative, so that the acceleration of a negative mass body takes place in the direction of the applied force, as is the case for a positive mass body.

16. When the mass experiencing a gravitational field is considered positive definite, while it is the direction of the gravitational field attributable to a given local matter distribution which itself varies under exchange of positive and negative energy observers, the same outcome as would occur when the equivalent gravitational field is reversed along with the mass of the body experiencing an invariant local gravitational field must be observed.

17. Not only is it allowed that the principle of relativity, which motivates the equivalence principle, be preserved by the proposed alternative concept of negative mass, but in fact it is this very principle that requires such a concept of negative mass according to which only the difference or the identity between the signs of mass of two bodies has a physical significance.

18. Only when local inhomogeneities in the matter distribution are not superposed for positive and negative energy matter can there be an effect of acceleration or rotation relative to those matter concentrations.

19. Given the unavoidable similarity of the large scale distributions of positive and negative energy matter, the phenomenon of inertia as an effect of acceleration relative to the large scale matter distribution can only occur if a body with a given mass sign gravitationally interacts solely with the large scale distribution of matter having the same sign of mass

as its own, because otherwise the effects attributable to positive and negative mass matter cancel out.

20. The generalization of the equivalence principle made necessary by the existence of negative mass matter implies the physical nature of the gravitational field as resulting from particle interactions despite the fact that it still allows a geometrical treatment of gravitation, because the metric properties of space and time are now themselves relatively defined properties which arise as a consequence of an equilibrium of local and inertial gravitational forces which depend on the sign of energy of the bodies experiencing them.

21. The equivalence principle must be generalized in such a way that it applies not merely locally but only for a single elementary particle with one mass or energy sign at once for which there would never be a difference between acceleration and a gravitational field.

22. From a gravitational viewpoint a void in a uniform positive energy matter distribution is not equivalent in general to a void in a spherical matter distribution of finite size and positive energy bodies present on the periphery a void in an unbounded matter distribution would actually experience a repulsive gravitational force as a consequence of the absence of gravitational attraction from the matter that is missing.

23. Birkhoff's theorem does not contradict the preceding conclusion, because it is valid only in a universe that is spherically symmetric around any point and for a homogeneous and isotropic matter distribution this condition is met only in the absence of a local void in a uniform matter distribution.

24. It is inappropriate to assume that when we are considering a spherical region of the universe the rest of the universe surrounding that region can be considered as a hollow sphere simply on the basis of the fact that according to the cosmological principle matter is distributed uniformly in all directions.

25. In the presence of a spherical void in an otherwise uniform matter distribution spherical symmetry exists only at the center of the void so that the presence of the void would necessarily alter the equilibrium of gravitational forces anywhere else inside (and to some extent also

outside) the void in such a way as to produce a force that would be the opposite of that which we would attribute to the presence of an equivalent additional quantity of matter with the same energy sign in place of the void.

26. The mistake involved in the traditional interpretation of Birkhoff's theorem consists in considering that the surrounding matter which could influence the particles located inside a chosen spherical region in a homogeneous and isotropic universe is spherically distributed around the center of the spherical region considered, instead of recognizing that the center of mass in a universe without boundary is always located at the position of the observer experiencing the effects of the matter distribution.

27. The gravitational repulsion that would be exerted on a positive energy body as a consequence of the presence of a void in a uniform positive energy matter distribution is actually the consequence of uncompensated gravitational attraction by matter with the same energy sign as that of the body.

28. Negative energy states are phenomenologically equivalent to an absence of positive energy from the vacuum, because removing positive energy from a vacuum with near zero energy is like decreasing energy into negative territory.

29. The expected gravitational repulsion exerted on a positive energy body by negative energy matter would occur as a consequence of the fact that the absence of positive energy from a region of the vacuum that is equivalent to the presence of negative energy matter would result in an uncompensated gravitational attraction from the surrounding positive energy vacuum pulling positive energy matter away from the region where the energy is missing.

30. A void in a uniform positive energy matter distribution remains physically distinct from a local absence of positive vacuum energy, even if in both cases the effects are equivalent to the presence of an excess of matter of negative energy sign, simply because an absence of matter (with positive energy sign) is necessarily different from the presence of matter (with negative energy sign).

31. Voids in the negative energy portion of the vacuum are equivalent to the presence of positive energy matter and along with voids in a uniform negative energy matter distribution would produce an equivalent gravitational repulsion on negative energy matter.

32. A void in a uniform negative energy matter distribution remains clearly distinct from a void in the uniform distribution of negative vacuum energy.

33. A description of matter of a given energy sign as voids in a filled distribution of *matter* of opposite energy sign would involve a violation of the requirement of relational definition of the sign of energy, because it would allow a forbidden absolute distinction between positive and negative energy matter, given that there can only be one filled matter distribution.

34. There must be a certain compensation between the usually considered contributions to vacuum energy which directly interact (other than through the gravitational interaction) only with positive energy matter and the usually ignored opposite contributions which directly interact only with negative energy matter, so that the natural value of the cosmological constant which we should expect to observe is actually zero.

35. It can no longer be assumed that there is a clear distinction between matter and vacuum given that matter is merely a manifestation of missing vacuum energy.

36. There can be no equivalent gravitational repulsion on positive energy matter from the presence of the void of cosmic proportion in the positive energy portion of the vacuum that is equivalent to a uniform distribution of negative energy matter.

37. The rate of universal expansion of matter with a given sign of energy is not influenced by the presence of matter with an opposite energy sign.

38. The voids in the negative energy portion of the vacuum which are equivalent to the presence of positive energy matter would interact with themselves even if the missing negative energy was uniformly distributed throughout all of space and despite the fact that a similar

distribution of missing positive vacuum energy would have no effect on positive energy matter.

39. The effect on a particle with a given energy sign of the presence of a void in a uniform matter distribution of opposite energy sign would be a gravitational attraction directed toward the void.

40. There can be no direct interactions of any kind between positive and negative energy particles, because no definite energy sign can be attributed to the fields of interaction between opposite energy particles and therefore negative energy matter must be dark.

41. Despite the absence of any direct interaction between opposite energy particles there exists an indirect gravitational repulsion between opposite energy bodies as a consequence of the equivalence between the presence of negative energy matter and a void in the positive energy portion of the vacuum which for positive energy bodies gives rise to an uncompensated gravitational attraction directed away from this void.

42. The absence of direct interaction between positive and negative energy particles does not mean that positive energy matter does not experience the gravitational effects of the negative energy portion of the vacuum, because as a particular manifestation of negative vacuum energy, positive energy matter cannot be expected not to interact with the negative energy portion of the vacuum.

43. Opposite action particles with opposite charges (as observed from the forward in time viewpoint) cannot annihilate one another under normal conditions, because there are no direct interactions between particles with opposite action signs, which means that they cannot come into contact with one another except under conditions where the opposite energies involved are extremely high and the spatial scale very short, in which case the indirect gravitational interaction they do experience is no longer negligible.

44. The creation of pairs of opposite action particles out of the vacuum is prevented from occurring under ordinary circumstances as a consequence of the weakness of the indirect gravitational interaction between such particles which requires very high (positive and negative) energy particles to be created, while any particle produced in such a

way is likely to immediately annihilate with an opposite action particle produced through similar processes taking place continuously in the vacuum.

45. Despite the fact that any process of pair creation involving opposite action particles which would occur in the vacuum would normally be followed by annihilation to nothing on a time scale characteristic of quantum gravitational phenomena, it is to be expected that such particle pairs could nevertheless be permanently created during the first instants of the Big Bang when the expansion of space takes place at a sufficiently high rate and this may actually explain the presence of matter in our universe.

46. A negative energy matter particle cannot decay to 'lower', more negative energies by emitting positive energy radiation particles, because the positive energy radiation particles could not even have been into contact with the decaying negative energy particle.

47. A positive energy matter particle cannot turn into a negative energy particle by emitting positive energy radiation particles, given that there can be no direct interaction between the now negative energy matter particle and the positive energy radiation it would have released.

48. No interaction vertex involving particles with mixed action signs needs to be taken into account in determining the transition probabilities of quantum processes.

49. For negative energy matter the objectively defined low direction on the energy scale, along which the thermodynamically favored degradation of energy occurs, is that toward the zero energy level of the vacuum as is the case for positive energy matter.

50. Negative energy matter particles do not have a natural tendency to 'decay' to states of larger negative energy through the absorption of negative energy radiation in the future direction of time for the same reason that positive energy matter particles do not naturally tend to reach states where a larger amount of energy becomes concentrated into fewer particles as a result of the absorption of positive energy radiation.

51. Based on the preceding results we can expect that there is no interference on the part of negative energy particles into the processes usually described by quantum field theory except at the energy level associated with quantum gravitational phenomena.

52. The energy that is gained or lost by a positive energy body as a consequence of its indirect gravitational interaction with a negative energy body is compensated by a variation in the negative gravitational potential energy associated with the variation of positive vacuum energy that is equivalent to the variation of energy of the negative energy body (a conclusion which is valid in the context where the negative gravitational potential energy associated with the interaction of this positive vacuum energy with the rest of the matter and energy in the universe naturally compensates the positive vacuum energy itself).

53. Given that the negative energy of a field of interaction between positive energy particles in a bound system cannot be directly and independently measured it cannot be assumed to contribute independently to the inertial mass of the entangled system as a whole.

54. The perpetual motion argument against gravitational repulsion only rules out the possibility of an anomalous gravitational interaction between ordinary matter and ordinary antimatter, because while a positive mass body could perhaps gain potential energy by being raised in the gravitational field of a positive mass planet by a negative mass body, this negative mass body would lose potential energy in the process, which means that no work can be produced in such a way.

55. Negative energy matter could not be used to provide the conditions necessary to make a traversable wormhole given that it cannot be made to remain near the singularity of a positive mass black hole or even simply be brought inside such a black hole.

56. The fact that negative energy matter cannot cross the event horizon of a black hole means that it cannot reduce the mass of the black hole and the area of its event horizon, so that the existence of negative energy matter would not allow to produce a diminution of the entropy associated with those objects.

57. Negative thermal energy from a negative energy system is not equivalent to negative heat for a positive energy system given that from the viewpoint of positive energy matter, kinetic energy is exchanged with negative energy particles as if it was a positive definite quantity and therefore the existence of indirect gravitational interactions between opposite energy systems does not allow the transformation of useless forms of thermal energy into more useful forms in a way that could have given rise to a reduction of entropy.

58. The observer dependence of the gravitational field which must be assumed in the context of a bi-metric general relativistic theory implies that observers with opposite energy signs experience the metric properties of space and time associated with a given local matter configuration in a different way.

59. There are two distinct categories of contributions to the total stress-energy of matter entering the generalized gravitational field equations that determines the metric properties of space experienced by a positive energy observer, the first is provided by the conventional stress-energy tensor and is positive definite for all densities of positive energy matter while the other is provided by the irregular stress-energy tensor and can be either positive or negative depending on the value of energy density of negative energy matter relative to its average density.

60. The natural value of positive and negative contributions to vacuum energy density is provided by the Planck scale and when positive energy is missing from the vacuum as a consequence of the presence of negative energy matter the energy of the vacuum is reduced locally from this maximum positive value.

61. In the proposed generalized gravitational field equations, the value of vacuum energy density observed by positive energy observers and associated with the cosmological constant is determined solely by the metric conversion factor associated with the map of the metric properties of space experienced by negative energy matter as negative energy observers measure them to those experienced by negative energy matter as positive energy observers measure them.

62. The values of vacuum energy density which are observed in the absence of matter by positive and negative energy observers could in principle

vary with position and time given that they arise as a consequence of applying the variable metric conversion factors which provide the map of the metric properties associated with matter of a given energy sign as they are experienced by observers of opposite energy sign.

63. Given that the observed value of the cosmological constant is very small it follows that there must be a near perfect level of symmetry at the present epoch between the metric properties of space experienced by positive energy observer on the cosmic scale and those experienced by negative energy observers.

64. Even when the cosmological constant is positive the generalized gravitational field equations describing the motion of negative energy matter are symmetric with the equations describing the motion of positive energy matter.

Chapter 2

Time Reversal

2.1 The problem of discrete symmetries

In this chapter I would like to explain how a more consistent and adequate formulation of the discrete P, T, and C symmetry operations involving a revised concept of time reversal can be obtained that integrates the insights gained while studying the problem of negative energy and that offers a better understanding of why and how such symmetries can under certain circumstances appear to be violated. Discrete symmetry operations are usually assumed to be relevant only in the context of quantum field theory, but in fact they can also be examined from a semi-classical standpoint. Their level of application is actually right at the interface between the classical world of gravitation theory and that of quantum theory and it should not come as a surprise therefore that some of the results which I have obtained will allow progress to be achieved concerning the problem of identifying the origin of the degrees of freedom associated with black hole entropy, which arises merely in a semi-classical context. In order to do so it will be necessary to introduce an additional category of discrete symmetry operations that relates positive and negative action matter particles in a way that is similar in many respects with that by which the charge conjugation symmetry operation relates ordinary matter and antimatter.

I had long ago realized that it would be necessary to revise our conception of space and time reversals, because the current formulation of those symmetry operations is based on unreasonable assumptions regarding the significance of time reversal and its relationship with the sign of energy and

that of non-gravitational charges. It is indeed presently believed that the charge conjugation or C symmetry operation is not a discrete space or time symmetry operation, but simply an additional symmetry having to do with charge as an independent concept. But I came to suspect that the relationships which are known to exist between this charge reversal operation and the discrete P and T symmetry operations associated with space and time reversals are an indication that C should be conceived and explicitly defined as a particular instance of discrete spacetime symmetry operation. What constitutes the underlying basis of those considerations is the acknowledgement that the sign of certain physical quantities (including charge) are dependent on their direction of propagation in time. From that viewpoint it would seem indeed that both the T and the C symmetry operations should be assumed to involve some form of time reversal and this is reason enough to suspect that they may also both give rise to a reversal of charge.

The problem, however, does not really have to do with the current conception of the charge reversal operation as such. What is truly inappropriate is the simple kinematic representation of time reversal as involving a backward motion of all particles and their angular momenta, which I believe is too rudimentary to characterize a reversal of the fundamental time direction degree of freedom. I also think that if T is to be assumed as actually reversing time then it should leave momentum unchanged (despite common expectations) as this is a quantity that should rather be reversed independently, along with the direction of space intervals. In this context if some reversal of momentum may still be of relevance to T it would clearly have to arise as a consequence of the fact that it is actually equivalent to the effects we should expect from an appropriate reversal of time when we insist on measuring physical quantities against the perceived rather than the actual direction of the flow of time. In any case it must be understood that what we observe from our classical historical perspective is not representative of the true evolution that takes place when we are dealing with the propagation of elementary particles. The subtleties of what is going on at the microscopic level are not directly apparent from the superficial viewpoint associated with a global representation of events 'after the fact' that provides a static picture of the spacetime paths followed by elementary particles. Therefore, it is not appropriate to define a reversal of the fundamental (non-thermodynamic) time direction degree of freedom based merely on narrative aspects of phenomena which are all directly discernible at this superficial level of description. Better formulations of the discrete spacetime symmetry operations are re-

quired which would reflect the actual and sometimes unrecognized variations or absence of variation of physical parameters associated with each of those reversals of the fundamental space and time direction degrees of freedom.

2.2 The constraint of relational description

To begin this discussion, I must first of all mention that once again the most significant constraint which we need to consider and against which our understanding of the discrete symmetry operations must be developed is that of the necessary relational definition of physical quantities and their changes. Those quantities are here the directions of space and time intervals, the directions of momentum and angular momentum and the signs of energy and non-gravitational charges. The main point I want to emphasize is that there can be no meaning in considering a change of any one of those quantities (to its opposite value) that does not occur relatively to some remaining unchanged parameter of the same kind. Breaking that rule is to be considered logically impossible simply because if it was allowed it would mean that we can define an absolute (metaphysical) direction or polarity (in the general sense), which in effect would not be related to any reference point of a physical nature in our universe. What I'm suggesting is that the profound reason why a certain level of lopsidedness, such as the observed breaking of P symmetry by the weak interaction, can exist is that such asymmetries merely occur when one or two physical parameters are reversed *relative* to a fixed background of unchanged directional parameters of a similar kind. In other words, what makes these violations of discrete symmetry possible is simply the fact that application of a reversal operation to a single parameter leaves some other properties unchanged which allows the asymmetry to occur as a real feature characterized by a measurable change relative to a distinct physical quantity. In the case of P symmetry, the reversal of space intervals involved occurs relative to the direction of time intervals which remain unchanged by such an operation and therefore it should be expected that violations of P can be observed given that the reversal of physical parameters associated with this operation can be measured against the unchanged properties.

But those asymmetries cannot imply the existence of an absolute lopsidedness or directionality at the most fundamental level for the universe as a whole, because they can be compensated by an appropriate reversal

of the unchanged parameters relative to which the original transformation took place. This is what explains that despite the violation of P symmetry by the weak interaction it remains impossible to provide an absolute definition of left and right, because indeed reversing the sign of charges allows to regain invariance. Thus, contrarily to what is sometimes assumed, the preferred handedness unveiled by the weak interaction is not more profound than that we observe in certain complex structures. As long as invariance under a more general discrete symmetry operation like CP is observed to hold, it is impossible to communicate the significance of right and left without knowing which of two C-related particles is to be considered as having positive electric charge. But if it is impossible to distinguish an absolute (non-relational) difference between positive and negative charges themselves, as I previously suggested, then only observers which are actually sharing the same universe and which are allowed to directly compare physical quantities, could differentiate between left and right.

This is a very general feature which I think would always be observed to apply given that it is actually required by the condition of relational definition which is relevant to any change of direction or polarity (such as a reversal of the sign of charges). The directions of space and time which are singled out by any process which appears to violate a discrete symmetry are significant only in relation to other aspects of reality which must be identifiable from within the universe in which those processes take place. If in one particular instance it was to be found that no combination of discrete symmetry operations allowed invariance to be regained, then it would mean that there exist physical properties which can refer to elements of reality not shared only by observers within our universe. In other words, if directional asymmetries not occurring merely in relation to unchanged quantities (not defined as mere relative properties) were allowed, it would in effect be impossible to describe the polarities so revealed by referring only to measurable properties of physical reality.

The problem which there would be if such violations of discrete symmetry were possible is that completeness and self-determination are the defining characteristics of the universe concept, in the sense that the universe is precisely that ensemble of physical elements which are all causally related to one another and to nothing else. Thus, if we were to find that the description of our universe can refer to absolute and immaterial notions of direction not defined merely as relationships between elements of reality which must be part of that universe, then the only logically valid conclusion would have to

be that there exists a *causally related* reality outside what we consider to be the universe (this has nothing to do with the concept of the multiverse whose elements are not to be assumed as causally related to one another) relative to which the otherwise metaphysical polarities could be properly defined. As a consequence, there is definitely no way our universe could be considered lopsided if it is actually the whole universe and I believe that the fact that it can be shown that the existence of such an irreducible asymmetry would imply that some physical quantities may not be conserved for the universe as a whole is a confirmation of the validity of this conclusion. It must be understood, however, that the identified requisite does not mean that symmetry could never be preserved following a reversal of one single parameter, like space direction alone, which can be defined in a relational way, but simply that such invariance is not absolutely required to apply under all circumstances.

Given those considerations, we can be totally confident that there is no such thing as an absolute direction of space or time intervals, because indeed this would imply a violation of the principle of relativity (as understood in its most general form which predates relativity theory) and the validity of this criterion is necessary for the consistency of any model concerning physical reality. Even without going into elaborate mathematical arguments, such as those entering the CPT theorem, it is therefore possible to appreciate that the only problem there could be in relation to the observation of an asymmetry under a properly defined discrete symmetry operation would have to involve a violation of invariance under a combined operation that reverses all parameters and leaves absolutely none unchanged. I will later explain why an appropriately defined PTC transformation must be considered as one instance of such a symmetry operation that reverses all parameters and leaves nothing unchanged (by actually reversing all space- and time-related parameters twice) and which we are thus justified to categorize as inviolable. But I believe that the fact that it would be impossible to provide a mathematical framework for quantum field theory that would satisfy the requirements set by special relativity if the equations of the theory are not invariant under PTC (which constitute the substance of the argument behind the traditional CPT theorem) confirms that relativistic imperatives (all measures of space and time intervals are relative) are the true constraints which impose invariance under the most general, combined, discrete symmetry operation.

The fact that this simple but most unavoidable requirement has never been considered as a means to restrict allowed violations of discrete sym-

metry illustrates the fact that our treatment of space and time reversals is incomplete and inadequate due to multiple misconceptions which do not concern only the aspect discussed here. The often met remarks to the effect that there is no a priori reason why the universe could not be asymmetric in a fundamental way and that it is only the above mentioned mathematical requirements arising from the CPT theorem that motivate the conclusion that some overall symmetry must nevertheless be obeyed under all circumstances are therefore inappropriate and misleading. But it should also not come as a surprise that the discrete symmetry operations, when performed independently from one another, may not produce invariance. What justified the unexpectedness of the violations of P and CP symmetries when they were first observed is actually the intuitive belief that absolute directionality should not be allowed, while, as I just explained, this is rather the argument that would apply to a more general symmetry operation like PTC whose required conservation, ironically, is usually not believed to be intuitively explainable. The truth is that, for an imbalance under reflection to exist, all that is required is that the world be unbalanced with respect to something. This conclusion is the outcome of the most unequivocal interpretation of the requirement of relational definition of physical quantities, which itself constitutes the one rule we can be most confident need to apply to the physical world we experience. In fact, the argument against the possibility of a violation of symmetry under a combined reversal of all space- and time-related parameters is probably the strongest kind of argument which can be proposed from a theoretical viewpoint.

Regarding time reversal in particular and the question of what it would mean to assume that the whole universe is running backward in time and whether there can be any objective meaning to such a reversal operation I think that given the preceding discussion we would have to recognize that such a reversal could in effect be physically significant if it is defined as a reversal that leaves other parameters, such as the direction of space intervals unchanged. But this means that such a time reversal operation cannot consist in a mere reversal of the motions and rotations of objects taking place in a reverse chronological order. A reversal of time that would be relationally defined would have to be meaningful both globally and locally as it would allow a distinction between a physical system with unchanged time direction and one with reversed time. This difference could be determined by directly comparing the physical properties of one of the systems with those of the other, if the two systems are part of the same universe. But a difference

could also be identified as occurring for the whole universe in relation to the unchanged direction of space intervals. In any case the above discussed constraint would require that such a relative backward in time evolution be clearly identifiable from the physical properties of the particles involved, precisely because it is only under such conditions that the change of direction in time could be objectively determined by comparing it with that of the unchanged parameters. But given that those differences would then actually be determined in relation to the value of parameters which are themselves reversible, it follows that no absolutely characterized notion of asymmetry would be involved.

In the context where absolute lopsidedness is to be considered impossible it follows that it is of primordial importance to identify all the physical properties which can be related to one another and which could be affected by transformations of the kind that involve a reversal of space and time directions at the fundamental level. Indeed, if we are to be able to determine whether there remain quantities not reversed when a certain discrete symmetry operation is performed, we certainly have to be able to determine which quantities are actually affected by the operation involved. It is my belief that some of the violations of discrete symmetries which are usually assumed to have been observationally confirmed are actually a consequence of the fact that the effect of the considered reversals on certain quantities are not taken into account, while invariance would actually be inferred if all quantities dependent on the parameters which are assumed to be reversed were appropriately transformed. I already mentioned the fact that there are indications to the effect that we may, in particular, expect the sign of charges to be dependent on the sign of time intervals experienced by the particles carrying them. Yet the traditional definition of the time reversal operation T does not involve any reversal of charges (from whatever viewpoint) and thus we could observe violations of such a T symmetry that would occur simply because we do not appropriately reverse the sign of charges when we try to verify invariance under a reversal of time (from a certain viewpoint). We must therefore first take care of identifying all unaccounted dependencies which may confuse our assessment of symmetry violations before we can truly appreciate under which conditions they are actually allowed to occur.

2.3 The concept of bidirectional time

Concerning the problem of discrete symmetries another essential aspect must be recognized in addition to that regarding the necessity of a relational definition of all such symmetry operations. Awareness of what it involves is of the highest importance for a proper resolution of all matters associated with time directionality and given that this is the central problem with which this report is concerned it is crucial to grasp the significance and the implications of the notions involved. Basically, what must be understood is that a distinction is to be made between the traditional concept of time direction associated with changes occurring at a statistically significant level where the notion of entropy is meaningful and a concept of time direction associated with the existence of a fundamental time direction degree of freedom independent from the constraints related to entropy variation. The traditional concept of time direction related to statistically significant changes and the growth of entropy gives rise to what I call the unidirectional or thermodynamic time viewpoint, while the alternative concept of time direction related to the existence of a fundamental time direction degree of freedom independent from statistical constraints gives rise to what I call the time-symmetric or bidirectional time viewpoint.

Associated with this alternative concept of time direction is a different notion of time reversal not limited by the constraints imposed on our description of physical processes by the second law of thermodynamics. Indeed, the traditional notion of time reversal associated with the thermodynamic time viewpoint merely consists in assuming a reversal of the motion of all particles involved in a process, so as to give rise to the same events as observed in the original process, but in the reverse order. However, those events would still be described from the same unique and immutable forward direction of time associated with entropy growth. This is a consequence of the fact that the unidirectional time viewpoint involves considering that there can only be one direction in time at once for the propagation of all particles, indiscriminately, which actually amounts to ignore the existence of a fundamental time direction degree of freedom. From that viewpoint if time was reversed all particles would have to propagate backward, not relative to some fundamental time direction parameter, but in comparison with the direction of motion which they were all *observed* to have originally. Thus, the time reverse of a process would simply be the equivalent process for which the same observations are made, but in the reverse order. The bidirectional or time-symmetric view-

point on the other hand is at once less restrictive and more distinctive in that it actually recognizes the existence of a fundamental time direction degree of freedom, distinct from the observed direction of motion of particles apparent to an observer constrained by the law of entropy increase. This time direction parameter must be allowed to vary from one particle to another, even between those of an otherwise identical nature which are involved in the same process at the same time.

Now of course I have already discussed the significance of the existence of a fundamental time direction degree of freedom as being that property which allows to explain the distinction that exists between a particle and its antiparticle, despite the fact that from an observational viewpoint both objects appear to be ordinary particles traveling forward in time, but which merely happen to carry opposite non-gravitational charges. However, I previously made clear that in fact the sign of charge is *not* affected by a reversal of the direction of propagation in time which may relate a particle with its antiparticle and therefore if it is nevertheless observed as being reversed it can only mean that the direction of time relative to which we measure the charge is not the true direction in which the particle is propagating in time, because an observer measuring the same physical property while following the true direction of propagation in time of the particle would not observe any change[1]. It is merely the fact that a backward in time observation is indeed impossible that justifies assuming a reversal of charges for a particle propagating toward the past. Indeed, measuring apparatuses always record changes as they occur in the future direction of time due to the fact that the processes involved in the amplification of the signal which gives rise to a measurement can only take place in this direction of time in a universe where a thermodynamic arrow of time governs the evolution of processes involving a large number of independently evolving particles. This constraint is therefore what justifies the use of a unidirectional viewpoint relative to which all physical properties are given as they would appear relative to the conventional future direction of time, even when the true direction of time in which the processes involved occur is the past direction. Non-gravitational charges,

---

[1]I will henceforth use the term 'propagation' in place of 'motion' to designate the true direction in which a particle is traversing space and time intervals, as occurs from a bidirectional time viewpoint. This allows to explicitly refer to those aspects associated with the fundamental time direction degree of freedom which are ignored from the viewpoint of unidirectional time relative to which all changes refer to a particle's observed (semi-classical) trajectory.

therefore, actually remain unchanged from the bidirectional viewpoint when the fundamental time direction degree of freedom is reversed, but this is the very reason why they appear to be reversed from the unidirectional time viewpoint.

A rule thus emerges which is that, for any particle propagating in the past direction of time, a time direction-dependent physical property of that particle which would be positive when considered from the bidirectional time viewpoint (relative to the true direction of propagation of that quantity in time) would appear as negative from the unidirectional time viewpoint. But this reversal of observed quantities from their true value is not restricted to charge or energy, which I had already identified as properties dependent on the direction of propagation in time, but would actually have to apply to the direction of space intervals associated with the motion of particles (which are always given in relation to time intervals) and thus also to momentum (even if the time intervals entering the traditional definition of momentum were assumed positive definite as a consequence of adopting a unidirectional time viewpoint). Thus, if momentum was assumed to be left unchanged by a properly defined reversal of time (on the basis of the fact that from a fundamental viewpoint the associated direction of space intervals is an independent parameter to be reversed by an independent symmetry operation, as I will later explain), it would nevertheless appear to be reversed in comparison with its actual value, from the unidirectional time viewpoint. But given that the direction of momentum is not fixed for a given type of particle propagating in a given direction of time (it also changes when the direction of propagation of the particle in *space* is reversed) it cannot be taken as a clear indicator of the direction of propagation in time of a particle. That, however, is not the case with charge, which from the bidirectional time viewpoint remains unchanged even as a particle reverses its direction of propagation in time (while also reversing its energy sign) and this is why it is possible, from the unidirectional time viewpoint, to identify the true (even if merely conventionally defined) direction of propagation in time of a particle based on the observed value of its non-gravitational charges (in relation to those of an otherwise identical particle)[2].

---

[2]In fact, even if this relationship between time direction and observable charge was valid only for ordinary particles and antiparticles, in the context where it would be possible to conceive of an independent operation of charge reversal that would reverse charge not merely from a unidirectional time viewpoint (as result of reversing the direction of propagation in time of particles) but even from a bidirectional time viewpoint, this con-

What is important to understand is this interdependence of space and time intervals even as they would be separately and independently transformed by their respective discrete symmetry operations. Thus, when we reverse the direction of the motion of a particle in space we reverse the sign of the space intervals associated with this motion not merely relative to the space axes, but also relative to time intervals (same time interval, opposite space interval). The sign of space intervals associated with the propagation of a particle submitted to a reversal of space directions would be reversed not merely from what it previously was (or relative to the space intervals associated with the motion of a particle not subject to the reversal), but also relative to the direction of time intervals in which the particle is still propagating. A particle which was propagating to the right relative to the future direction of time will now be propagating to the left relative to the same future direction of time, which was not affected by the reversal of space directions (this is illustrated in figure 2.1 where I consider the effects of the various discrete symmetry operations as they will be defined below). In other words, the particle is not just propagating left, it is propagating left forward in time, because indeed we are always concerned with the properties of processes involving particles propagating in space and time and not just with the properties of space or time themselves. What matters therefore is not just the direction of space intervals associated with some arbitrarily fixed spatial coordinate system, but the direction of space intervals for a particle propagating in a given direction of time, as asserted from a fundamental bidirectional viewpoint. Similarly, when time is assumed to be reversed it must be considered that the time intervals are reversed relative to the unchanged direction of space intervals in which a particle submitted to the reversal is propagating, so that the same positive space intervals are now traveled in the opposite direction of time. This does not mean that a reversal of both space and time cannot have clear meaning, however, because as I will explain later, even in such a case there would still remain unchanged physical properties relative to which the transformation could be characterized.

This relationship between space and time intervals is what gives a true physical meaning to the notion of time reversal when it is to be considered as a symmetry operation clearly distinct from space reversal and which should

---

clusion would still be valid, because as I will explain in section 3.3 particles carrying such a reversed charge would remain clearly distinguishable from ordinary particles and antiparticles, regardless of the direction of time in which they are propagating.
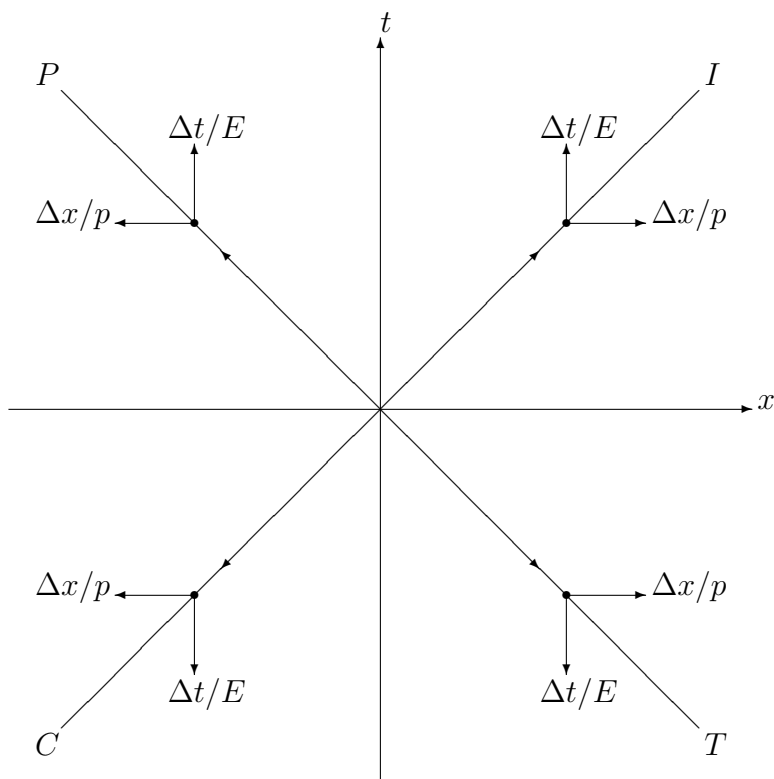
Figure 2.1: Variation of physical parameters under the proposed alternative definition of P, T, and C as described from the bidirectional time viewpoint. In this figure and the other related figures, I represents the original state and the diagonal lines correspond to particle trajectories. The space and time intervals Δx and Δt are indicated by vectors whose lengths correspond to the magnitude of the intervals and whose directions indicate the sign of the intervals relative to the space and time coordinates. The direction of the vectors associated with the energy E of particles corresponds with the sign of energy relative to the direction of the time coordinate.

therefore leave momentum unaffected (from the bidirectional viewpoint at least). In fact, it is what allows the very notion of a fundamental degree of freedom associated with direction in time to have a definite meaning, because it allows to distinguish (as a theoretical possibility) the process by which a particle is going through a given spacetime trajectory forward in time from the similar process by which an identical particle would be going through the exact same spacetime trajectory, only now backward in time. Such a distinction is crucial given that if we were to ignore it then from a unidirectional viewpoint in time there would be no meaning to assume that it may be possible for a trajectory to be traversed backward in time, given that from such a viewpoint we always observe particles as if they were necessarily going forward in time. But given that charge can be assumed to be left unchanged by a reversal of time (from the bidirectional viewpoint) we are actually allowed to differentiate between those two situations from an observational viewpoint, even in the context where all particle trajectories are necessarily followed as if they were occurring in the 'normal' chronological order (forward in time) associated with the growth of entropy, regardless of the true direction of propagation in time of the particles. It is therefore the relation between space intervals and time intervals that allows to distinguish backward in time propagation from forward in time propagation and the fact that the observed value of the sign of charge is dependent on that distinction simply confirms that it is appropriate to consider the existence of such a directionality parameter for the time dimension at the fundamental, elementary particle level.

It must be clear, however, that the coordinate systems for space and time still have a physical significance, because you may reverse the direction of the space intervals traveled by particles in the forward direction of time as well as the associated momenta while keeping the positions of the particles in space unchanged (not reversed as they would under a conventional space reversal operation). Indeed, as a comparison of figures 2.1 and 2.2 allows to reveal, it is only from the bidirectional time viewpoint that the sign of space and time *intervals* corresponding to the directions of propagation of particles always change in association with the sign of *positions* on the space and time coordinate axes, while from the unidirectional time viewpoint that need not be the case. Under such conditions quantities like angular momentum, which depend on both the position in space and the direction of space intervals, may not always be left invariant as they would when a complete space reversal operation is performed. This would occur in effect for processes submitted to

a reversal of time when they are described from the unidirectional viewpoint in which time is maintained positive even for backward in time propagating particles and all time direction-dependent quantities like the direction of space intervals and the momentum of a particle consequently appear to be reversed, while the positions are left unchanged (which implies that spin would appear to be reversed). In this context it seems that space intervals, as properties defined in relation to the direction of propagation in time, can actually be reversed in two different ways. They may be reversed because space directions are reversed (which also reverses positions) or they may be reversed because the direction in which they are assumed to be traversed in time is reversed (which leaves positions unchanged). This distinction is what allows the traditional concept of time reversal as affecting the directions of momentum and angular momentum to still be relevant, even in the context of the existence of a fundamental time degree of freedom, when those directions should in fact be left invariant (from a bidirectional viewpoint) by a properly defined time reversal operation.

Another point must be emphasized regarding the kind of time reversal operation which can be developed in the above described context. Indeed, if we no longer consider appropriate the picture of time reversal as consisting in a simple reversal of the observed motion of each and every particle then it must also be recognized that a properly defined time reversal operation could never give rise to a reversal of the thermodynamic arrow of time for the physical systems involved. In fact, I think that we should already suspect that there is something wrong with the often met suggestion that a reversal of the motion of every particle in a region of space would give rise to entropy decreasing evolution (in the absence of any external perturbation). For such a proposal to be valid it would have to be shown that the origin of the observed time asymmetry of thermodynamic processes in our universe is to be found in a very precise adjustment of the motion of every single particle in the universe at the present time which would occur in just such a way as to allow a state of minimum entropy to be reached as time unfolds in the past right back to the Big Bang state.

However, given the inherently random nature of quantum processes and the extreme sensitivity to initial conditions (here the 'final' conditions giving rise to a given past evolution) which are known to exist even in a classical context, this hypothesis appears highly implausible (I will address this question more thoroughly in section 3.6). But if in addition we admit the existence of a fundamental time direction degree of freedom distinct from the
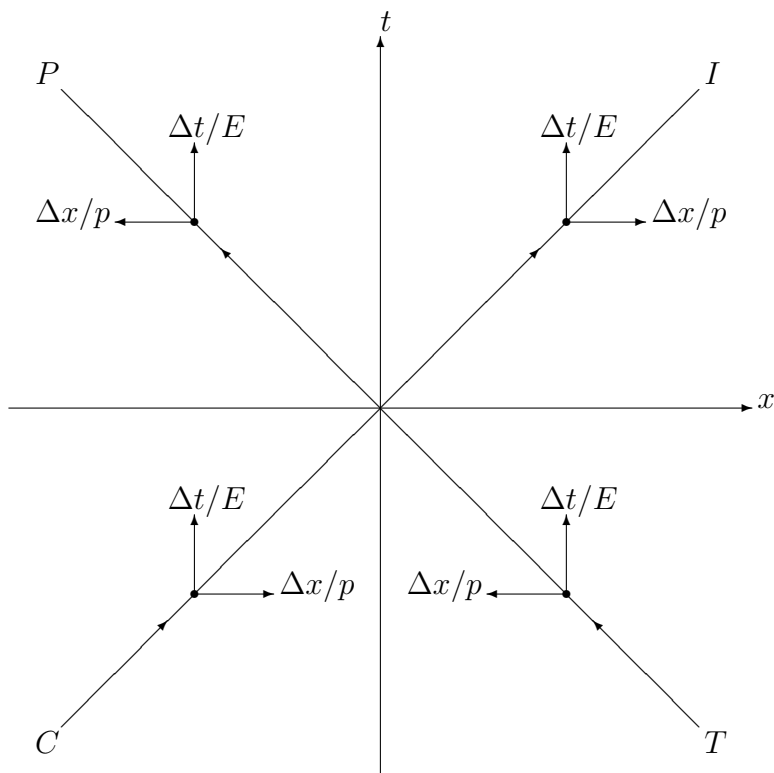
Figure 2.2: Variation of physical parameters under the proposed alternative definition of P, T, and C as apparent from the unidirectional time viewpoint. We can see that from this viewpoint the only difference between the original process and the T-reversed process is that the space intervals are traversed in the opposite direction, just as would be expected according to the traditional definition of backward in time motion. The case of the C-reversed process is also quite in line with traditional expectations given that such a process should not be different from the original process except for a reversal of the sign of charges (which is not illustrated here) which would in fact also occur for the T-reversed process despite traditional expectations.

observed motion of particles then we clearly have to reject the possibility that a reversal of time may produce anti-thermodynamic behavior, because time-reversed propagation is in fact already taking place in processes for which there is no apparent change to the direction of the thermodynamic arrow of time. This means that the direction of propagation in time of particles (the sign of time intervals associated with a bidirectional viewpoint) is not necessarily that relative to which entropy increases despite the fact that it may appear unnatural that evolution could proceed in a direction of time other than that in which we do observe time to be 'flowing' (as a thermodynamic necessity). The thermodynamic arrow of time and the notion of time directionality occurring from a bidirectional viewpoint are two completely independent concepts.

2.4   Alternative definition of C, P, and T

One last remark is necessary before I can provide a full description of exactly how the fundamental physical properties of matter should be considered to vary under an alternative set of discrete symmetry operations formulated so as to allow the above discussed requirements to be satisfied. I previously hinted at the fact that the direction of momentum should be considered as independent from the direction of time at least from the most consistent viewpoint which is provided by a bidirectional perspective on time. I believe in effect that momentum, as the attribute conjugate to physical space, should only be considered to reverse along with space and not along with time, just as energy being the physical attribute conjugate to time should necessarily reverse when time reverses and only then. There is, however, an additional motivation for requiring this kind of joint variation of all space-related attributes or time-related attributes (independently) besides the fact that consistency may require that it be imposed when what we seek to assert is precisely the dependence of various parameters under reversal operations which are defined after the quantities they are assumed to reverse. This perhaps more unavoidable justification for the joint variation of conjugate attributes is to be found in the requirement that the considered symmetry operations should not change the sign of action of the physical systems on which they operate.

It is my understanding of the true physical significance of a reversal of the sign of action that allows me to recognize the necessity to define the discrete

symmetry operations in such a way that momentum would necessarily reverse as a consequence of a reversal of space coordinates while energy would necessarily reverse as a consequence of a reversal of the time coordinate. Indeed, in the context where a reversal of space coordinates would necessarily give rise to a reversal of space intervals, while a reversal of the time coordinate would necessarily give rise to a reversal of time intervals, if the sign of action itself is to remain invariant then it means that a reversal of space must also involve a reversal of momentum and a reversal of time must also involve a reversal of energy. In fact, we always implicitly assume that the P, T, and C reversal operations do not relate physical processes in which the particles involved would have opposite action signs or energies (as measured from the forward direction of time). But the implications this should have for the dependence (under conventional discrete symmetry operations) of the signs of momentum and energy on those of space and time intervals is not always recognized. I believe that this lack of clarity is responsible for a good part of the misunderstanding regarding what parameters should really be affected by any symmetry operation involving a reversal of time. In tables 2.1, 2.2, 2.3, and 2.4 I will therefore provide an explicit account of the dependence of the signs of momentum and energy, along with those of space and time intervals, under all relevant discrete symmetry operations. It will be apparent from this account that clear distinctions exist between the traditional and the redefined time reversal and charge conjugation symmetry operations. Yet, given that the original definitions actually need to be replaced and cannot even be considered meaningful anymore, I think that it will not be necessary to relabel those operations and associate them with new symbols or letters, so that I will continue to use the T and C notation when referring to those redefined discrete symmetry operations.

In the following tables and in the corresponding diagrams (figure 2.1 corresponds to table 2.3 and the bidirectional viewpoint, while figure 2.2 corresponds to table 2.4 and the unidirectional viewpoint) the position along the space and time axes are denoted x and t (I'm assuming a one-dimensional space for simplicity) while the space and time intervals corresponding to the motion, or the propagation of the particles involved in the processes which are transformed by the symmetry operations are denoted Δx and Δt respectively. The energy of the particles involved in the same processes is denoted E and can actually vary in sign, while the momentum of those particles along the x axis is simply denoted p. The sign of non-gravitational charges (which allows to distinguish between the state of a particle and that

Trad.	t	Δt	E	x	Δx	p	q	s	h
I	t	?	E	x	?	p	q	s	h
P	t	?	E	$-x$	?	$-p$	q	s	$-h$
T	$-t$	?	E	x	?	$-p$	q	$-s$	h
C	t	?	E	x	?	p	$-q$	s	h

Table 2.1: Variation of the physical parameters associated with a process transformed by the discrete P, T, and C symmetry operations as they are traditionally defined. The absence of explicit assumption concerning the Δt and Δx parameters (specifically) can be noted. The variation of the direction of angular momentum s as well as that of the handedness h can be derived from those of the other fundamental parameters, but the outcomes are nevertheless indicated here and in the other tables, because in certain cases they differ from what is traditionally expected. The identity operation I which corresponds to an absence of reversal is shown for reference purpose.

of its antimatter counterpart), even though it should be understood not to be reversed by any of the conventional discrete symmetry operations (including C) from the bidirectional time viewpoint (which provides the most accurate description of the transformations involved), is nevertheless denoted as q and may actually appear to be reversed from the unidirectional viewpoint. The sign of angular momentum related to the motion of the particles involved in the processes transformed by the P, T, and C operations, as well as the spin direction of elementary particles, which again should be understood not to be affected by those operations from a bidirectional time viewpoint are together denoted by the letter s, while the associated parameter of handedness (the direction of spin along the axis associated with the momentum of a particle) is here denoted h and should be expected to vary, even from a bidirectional time viewpoint.

From a semi-classical viewpoint, the displayed tables giving the variations of the time-related and space-related physical parameters under the traditional or redefined discrete symmetry operations, along with the assumptions which are made concerning the variation of the sign of charge, provide the most precise definitions that can be achieved of the operations involved. Using those definitions one can rebuild the quantum operators which are needed to transform the state vectors or the propagators corresponding to specific

Impl.	t	Δt	E	x	Δx	p	q	s	h
I	t	Δt	E	x	Δx	p	q	s	h
P	t	Δt	E	$-x$	$-\Delta x$	$-p$	q	s	$-h$
T	$-t$	Δt	E	x	$-\Delta x$	$-p$	q	$-s$	h
C	t	Δt	E	x	Δx	p	$-q$	s	h

Table 2.2: Implicitly assumed variation of physical parameters under the discrete P, T, and C symmetry operations as they are traditionally defined. The parameters whose transformation is only implicitly assumed are the space and time intervals Δx and Δt associated with the propagation of the particles involved in the processes transformed by the various discrete symmetry operations. The absence of reversal of Δt when time is assumed to be reversed can be noted.

Bidir.	t	Δt	E	x	Δx	p	q	s	h
I	t	Δt	E	x	Δx	p	q	s	h
P	t	Δt	E	$-x$	$-\Delta x$	$-p$	q	s	$-h$
T	$-t$	$-\Delta t$	$-E$	x	Δx	p	q	s	h
C	$-t$	$-\Delta t$	$-E$	$-x$	$-\Delta x$	$-p$	q	s	$-h$

Table 2.3: Variation of physical parameters under the redefined discrete P, T, and C symmetry operations as described from the bidirectional time viewpoint. The necessary reversal of Δt with E as well as that of Δx with p can be noted, as also the necessary reversal of t with Δt and that of x with Δx. This is the variation of physical parameters which would be produced by the most appropriately defined discrete symmetry operations that can be formulated in a semi-classical context. Here all reversals of physical quantities are seen to occur twice or to not occur at all, as required for explicit invariance under a joint PTC operation.

Unidir.	t	Δt	E	x	Δx	p	q	s	h
I	t	Δt	E	x	Δx	p	q	s	h
P	t	Δt	E	$-x$	$-\Delta x$	$-p$	q	s	$-h$
T	$-t$	Δt	E	x	$-\Delta x$	$-p$	$-q$	$-s$	h
C	$-t$	Δt	E	$-x$	Δx	p	$-q$	$-s$	$-h$

Table 2.4: Variation of physical parameters under the redefined discrete P, T, and C symmetry operations as described from the unidirectional time viewpoint. Again all quantities are reversed either twice or never by a combination of all operations, which guarantees explicit invariance under PTC. The equivalent reversal of charge q by both T and C as well as the apparent absence of any variation of Δt and E and the absence of joint variation of x and Δx when t is reversed can be noted.

quantum states or processes. It must be clear that quantum field theory itself does not dictate how the discrete symmetry operations should be defined and it is merely the assumptions used while formulating the related operators (to achieve transformations that match our expectations regarding which parameters should be affected by a given operation) that provide the necessary constraints on which depend their precise mathematical formulation. What I bring to the table, therefore, is an improved knowledge of the constraints that must apply to those transformations, based on a re-examination of the meaning of space and time reversals as they would occur in a semi-classical context. It is important to recognize indeed that despite the apparent freedom, the discrete symmetry operations cannot be arbitrarily defined, but must be the outcome of the most unavoidable consistency requirements (formulated in an empirically motivated context) which I believe are those I have identified in the above discussion. The fact that greater simplicity has been achieved while redefining those symmetry operations is only a further confirmation of the appropriateness of the alternative viewpoint that emerged from the preceding analysis. Indeed, the pattern of variations of physical parameters which is illustrated in figure 2.1 is strikingly simple in comparison with that we would have according to the traditional definition of the discrete symmetry operations and this simplification was actually one of the objectives I sought to achieve while redefining them. Let me then describe what the elegance of this proposal really embodies.

Looking at the tables in which the outcomes of the various discrete sym-

metry operations are displayed one thing we may first remark is that the parity operation P remains as it was originally defined, even in the context of the proposed alternative formulation of those transformations and this regardless of whether we use the bidirectional or the unidirectional time viewpoint. Of course the reversal of space intervals associated with the propagation of particles (which from my viewpoint must occur as a result of the reversal of space coordinates) is now more explicitly stated, but otherwise the traditional definition of space reversal remains unchanged. There is one good reason for that, which is that the revision I'm operating regards the concept of time direction essentially and the P operation is unique for being the only one that does not involve any time reversal, regardless of the approach favored. This is what explains that this operation was properly defined already, in the form it originally was, despite the failure of the traditional viewpoint in general. What P expresses indeed is a reversal of space coordinates that produces a reversal of positions, space intervals and naturally also momentum (as a requirement of action sign invariance) while it leaves unchanged (now as a matter of definition) the position in time, the time intervals and the sign of energy. No reversal of charge is to be observed in this case (particles are not replaced by antiparticles), from any perspective, because there is no time reversal involved from a bidirectional viewpoint and thus no change to be associated with the adoption of a unidirectional time viewpoint. There is no reversal of angular momentum either (because both momentum and position are together reversed), which is appropriate given that if angular momentum or spin were reversed a forbidden reversal of action would occur from the bidirectional viewpoint (because spin has the dimension of an action) that would not be associated merely with the shift to a unidirectional time viewpoint. But again this is in perfect agreement with traditional expectations regarding the effects of P. Handedness is to be assumed reversed by such a reversal of space, however, because momentum is reversed while spin is left invariant from all viewpoints.

It should be noted that the explicit mention of a reversal of space intervals Δx under a symmetry operation like P does not mean that a reversal of space intervals must be assumed to occur in addition to that produced by the reversal of space coordinates. In other words, if the space intervals are indeed reversed it is merely as a consequence of the reversal of space coordinates, as otherwise there would be no real change in the direction of space intervals, that is, no change relative to the new coordinates. We may in fact consider it more appropriate to assume that it is the intervals

themselves which are reversed along with the position of particles while the coordinates remain unchanged, which would still be equivalent to reversing the coordinates themselves. If I choose to explicitly mention a reversal of space intervals, along with the assumed reversal of positions, it is because there may be situations where the intervals would be reversed independently from the positions on the coordinate axes and we must be able to distinguish between the two situations. What the explicit statement of a reversal of Δx should be understood to imply, therefore, is that there must occur a reversal of the sign of space intervals traversed by the particles involved in the reversed processes in comparison with the sign of space intervals experienced by particles involved in processes which would not be submitted to the reversal. Those space intervals, therefore, are those which are traversed during unchanged time intervals and which we may ordinarily associate with the directions of the momenta of the particles involved. Indeed, the reversal of space intervals associated with the motion of particles is usually assumed to be implied by the reversal of momentum itself, but given that I will later suggest that momentum can be reversed without space intervals being equally reversed (when action is to be considered reversed) then it becomes necessary to explicitly define the variation of space intervals under P and to recognize that momentum direction is an independent quantity whose specification is not sufficient to determine the sign of space intervals spanned during a given time interval (except if action sign is in effect required to be invariant).

It must be recognized therefore that the reversal of Δx is not *merely* a reflection of the reversal of space coordinates, but is also a manifestation of the physical changes that occur when a particle reverses its direction of propagation in space while retaining its direction of propagation in time and those changes would be significant even if the position in space was to itself remain unchanged. Likewise, what the specific statement about the reversal of momentum p under space reversal P is intended to mean is that the direction of momentum is now the opposite of what it was, not merely relative to the new coordinates, but also relative to the directions of the momenta of particles which would not be subject to the symmetry operation. I may add that the same remarks would apply to time intervals Δt and the sign of energy, because if the reversal of those physical parameters under the T and C operations (from a bidirectional viewpoint) can be understood to occur as a consequence of the reversal of the time coordinate, it is clear that it also arises in relation to the time intervals experienced by particles which would be left unaffected by the reversal.

2.5   The time reversal operation

Despite a concordance of the rules for deriving the variation of physical parameters under any one of the redefined discrete symmetry operations there are important differences between the case of time reversal T or charge conjugation C and that of space reversal P and this is reflected in the fact that those two symmetry operations would produce results which are unexpected from a traditional viewpoint. In the case of T it must be required in effect that the physical time intervals Δt associated with the propagation of elementary particles and the energy E be together reversed when the time coordinate is reversed (if action is to remain positive when it already is), while it is traditionally assumed (even if only implicitly) that both energy signs and *bidirectional* time intervals are in fact unchanged by T despite the reversal of the time coordinate. Also, it must now be assumed that there is no *a priori* reversal of the space intervals Δx and momentum p when time is reversed (which is allowed when those parameters are recognized as independent from the time-related parameters Δt and E). This is required despite the fact that traditionally momentum is assumed to be dependent on time intervals (I will explain below how this apparent contradiction is to be resolved). In fact, the traditional assumption that p would be reversed by T, while the position x on the space axis would remain unchanged, would be problematic if in this context we did not again implicitly assume an independent reversal of physical space intervals Δx by presuming an invariance of the sign of action.

What must be recognized therefore is that from a consistent bidirectional viewpoint, when the time coordinate is reversed it must be assumed that the time intervals of propagating particles (associated with the fundamental time-direction degree of freedom) along with their energies (as defined relative to the true direction of propagation in time) are reversed, while momentum and space intervals are left unchanged, just like a reversal of space coordinates is assumed to imply a reversal of the space intervals and momenta, but no change to energy sign and no reversal of time intervals. This independence of space- and time-related physical parameters (from one another) is a requirement of the constraint of relational definition of those quantities which imposes that something remains unchanged when T or P is applied and those invariant properties are in fact the spatial directions themselves (when the direction of time is reversed) or the direction of time itself (when space directions are reversed).

Now, if we appropriately assume that the spatial positions, the space intervals, and the momenta remain unchanged under a properly defined time reversal operation it follows that the spin and the handedness must also remain invariant. Those relationships may appear unnatural (spin is usually considered to be reversed under a reversal of time), but from a bidirectional time viewpoint they are perfectly acceptable and in the context where we want to define time reversal as really affecting time-related parameters in a specific way, they actually constitute unavoidable requirements. What's more, the discussed invariance is derived from the bidirectional time viewpoint according to which the values of physical properties are such as they would appear to an observer following the direction of propagation in time of the particles involved in the processes submitted to this reversal. But from a unidirectional time viewpoint (of the kind that is required from a practical perspective) the only quantities which would appear to be left unchanged when time is reversed would actually be the time intervals Δt and the energies E, because they would be submitted to twice the same reversal, once as time-related quantities and once as a consequence of the additional reversal occurring when we are forcing a forward in time perspective. This is what justifies the validity of the assumption that energy would not appear to be reversed from the conventional forward in time viewpoint and it means that if energy was not in effect reversed from the time-symmetric viewpoint, then from the unidirectional viewpoint it would actually appear to be reversed by T, which is certainly not desirable.

On the other hand, the physical space intervals and the momenta associated with the propagation of particles do need to be reversed (once) when time is reversed if we insist on describing the motion of particles as it appears to take place from the conventional forward in time viewpoint and this despite the fact that only the physical time intervals experienced by the particles should actually be reversed by T. Indeed, given that the direction of space intervals is defined in relation to the direction of time intervals, if time intervals are followed in the wrong direction, then space intervals are also traversed in the wrong direction, so that the observed directions of the motion of particles are opposite the true directions of their motion, which means that those directions are actually reversed under a properly defined T operation when the outcome of this operation is considered from a unidirectional time viewpoint (this is made apparent when we reverse the direction of the arrows associated with the time reversed states in figure 2.1 to produce those in figure 2.2). Thus, when the direction of time is reversed, but

the time intervals in which the particles propagate are kept unchanged as a consequence of practical limitations imposed by the thermodynamic nature of the observation process, the associated space intervals actually appear to be reversed (they are the negative of those really experienced by the particles) even though the spatial positions remain unchanged. This is true again despite the fact that at the most fundamental level of description, which is that of bidirectional time, the direction of space intervals is to be considered unchanged by a reversal of time. As a consequence, we obtain results which comply with the traditional definition of time reversal according to which momentum (and implicitly also space intervals) should in effect be reversed by T along with angular momentum or spin, because given that momentum is here reversed independently from the position parameter x it follows that angular momentum would also appear to be reversed.

From the unidirectional viewpoint it may in effect seem like the traditional conception of time reversal as involving a reversal of motion which simply allows the particles to follow a trajectory backward could be valid. We must recognize, however, that just as there is no reason to assume that momentum is affected by a reversal of time from a bidirectional viewpoint (which explains that it is reversed from a unidirectional viewpoint), there is also no reason to assume that the sign of charge, as distinct from that of energy (the gravitational charge), would be affected from this same viewpoint when T is applied, because charge is not constrained to reverse by the requirement of action sign invariance when the direction of propagation in time reverses. This may also appear to comply with traditional expectations, but in fact (as I previously remarked) it rather constitutes the one aspect which introduces a radical departure from what is normally assumed concerning time reversal. Indeed, it means that the same reversal that does apply to momentum from the unidirectional time viewpoint would have to apply to non-gravitational charges as well, because if the direction of propagation in time of the charges is actually reversed as required, then the fact that time is followed in the same forward direction relative to which the charges were originally propagating means that the charges would now appear to be reversed. We must therefore consider a reversal of charges to be associated with a reversal of time, as a result of the fact that this physical property is not experienced along the true direction of time in which it is propagated. This is a very important result which is definitely not expected from a traditional viewpoint given that it asserts that a quantity which was previously assumed to be unaffected by a reversal of time (namely the sign of charge)

would actually appear to be reversed under such a transformation and if the preceding argument is valid then this conclusion would have to be considered unavoidable.

Thus, it seems that considering a reversal of time without assuming a consequent reversal of charge is incorrect and may give rise to violations of symmetry which are a simple artifact of the inappropriateness of traditional assumptions concerning which quantities are reversed along with the time coordinates, from the unidirectional viewpoint. To be meaningful, the experiments which seek to verify invariance under T would actually have to assume a reversal of momentum and spin retracing a process backward, but combined with a reversal of charge (a permutation of particle and antiparticle). In other words, to test the invariance of physical laws under time reversal we would have to use antimatter, which may explain why a violation of T symmetry is so difficult to observe despite the fact that violations of the combined CP symmetry were actually observed (which implies that T should also be violated given that CPT is inviolable). It appears that we are simply not using the right kind of matter to probe for T violation. It is not the invariance of a process relative to the thermodynamic arrow of time which must be probed, but invariance under a reversal of the true directions of propagation in time of elementary particles. I believe that the improved consistency of the interpretation suggested here from both an observational and a theoretical viewpoint confirms that the traditional definition of time reversal as involving nothing more than a reversal of the directions of motion and rotation of particles can no longer be considered appropriate.

It may also be noted that from a unidirectional viewpoint the reversal of charge and the reversal of spin under a properly defined time reversal operation are now the only aspects that differentiate this T operation from the P operation, apart from the respective reversals of the time and space coordinates themselves. But given that spin can also vary independently from the direction of propagation in time of a particle this means that the only unmistakable distinction between the time-reverse of a given state and the space-reverse of the same state is in effect the sign of charge, which again emphasizes the importance of recognizing the dependence of this parameter on the direction of time. In such a context it seems possible that the violations of T which may have been observed despite all the previously mentioned experimental difficulties could actually be violations of P symmetry, or violations of combined symmetries under which charge is left invariant by being reversed twice, because indeed those experiments do not compare matter

and antimatter processes. Yet it might be considered that, despite what is commonly believed, violations of time reversal symmetry had already been observed, even before the violations of traditional T symmetry were reported, because, as I will explain below, the C operation also involves some time reversal and violations of charge conjugation symmetry do occur. In any case it is clear that a violation of the time reversal symmetry operation T as it was here redefined would not provide us with an absolute direction of time at a fundamental level, but merely with a preferred direction of time relative to some arbitrarily chosen direction in space, or relative to some arbitrarily chosen sign of charge.

Another particularity of the alternative definition of time reversal proposed here is that it implies that it would now be electric fields which would reverse under application of the T operation instead of magnetic fields, because electric fields depend only on the sign of charge of the source particles and charge must be assumed to reverse under time reversal. Magnetic fields on the other hand would now remain unchanged under time reversal, because from the unidirectional viewpoint the direction of motion of the source particles would reverse, as is currently understood, but charge would also reverse, despite what is currently assumed, so that currents (which are the source of magnetic fields) would remain unchanged as a consequence of being submitted to this additional reversal. We must therefore assume that a relative change between the direction of an electric field and that of a magnetic field does in effect take place under a properly defined time reversal operation, only it is not attributable to a variation of the magnetic field, but rather to a variation of the electric field. The failure to recognize the dependence of the sign of charge on the direction of propagation in time of elementary particles therefore gives rise to an incorrect appraisal of the response of electromagnetic fields to a reversal of time.

A more consistent definition of the operation of time reversal on the other hand allows to avoid the troubling conclusion that certain phenomena involving electromagnetic fields would actually constitute a challenge to the necessary relational definition of discrete symmetry operations. Indeed, violations of time symmetry could arise for example in the case where neutrons would be observed to have an electric dipole moment and as such could effect a movement of precession around the direction of an external electric field, because this movement would appear to vary depending on the direction of time, but independently from the direction of the field and the sign of the electric dipole. However, while the direction of the dipole is not affected

by the reversal of a neutron's spin angular momentum occurring as a consequence of the reversal of time, according to my proposal it would nevertheless be reversed together with it, because it depends on the sign of the constituent particles' electrical charges, which we must now also assume to be reversed as a consequence of applying the T operation. It is not possible in this context to assume that a reversal of time would allow a change in the precession motion of the neutron (associated with the direction of the neutron's spin) to occur *independently* from the direction of its electrical dipole in the presence of an invariant external electric field, because in fact both the spin and the dipole must be assumed to be reversed by T, along with the external electric field. In other words, it is no longer possible to assume that while we should observe the precession motion to occur in reverse upon reversing time, the same dipole would nevertheless be interacting with the same electric field, as would happen if applying T actually reversed spin, but left the direction of the dipole and the external electric field unchanged. When the appropriate time reversal symmetry operation is considered, only *relative* differences can occur between the direction associated with the precession motion and the direction of the dipole.

Still concerning the T operation, it must be clear that it is not possible to assume that what the traditional definition of this transformation involves is a reversal of the time coordinate that reverses physical time intervals and leaves energy unchanged, combined with a reversal of momentum that leaves both space coordinates and physical space intervals unchanged, even if that would appear to correspond with the explicit definition of T as it is usually conceived. Such a definition of time reversal would be inapplicable simply because it would reverse the sign of action of the physical systems involved and this is certainly not desirable knowing that negative action matter (propagating positive energies backward in time) would be an entirely different kind of matter from a gravitational viewpoint and therefore certainly cannot be involved in those processes which we currently assume to be the time-reverse of processes involving positive action matter. This has nothing to do with the fact that a unidirectional viewpoint is used traditionally. It is a different problem that would be unique to the T operation despite the fact that I'm here assuming that C also involves some time reversal, because charge conjugation is simply not assumed to involve any space or time reversal traditionally and as such cannot be mistaken to involve action sign reversal. From the viewpoint of unidirectional time we can therefore only assume that the space intervals are reversed by T, along with the momenta, and that the time

intervals, along with the energies, are left unchanged by the same operation despite the reversal of the time coordinate. In other words, an appropriate (action sign preserving) time reversal operation needs to reverse both momentum and space intervals together (from a unidirectional viewpoint) or leave them unchanged together (from the time-symmetric viewpoint) and those constraints must be explicitly stated in the definition of the symmetry operation. This again illustrates how important it is to identify the variability of all physical parameters under any discrete symmetry operation, in particular for what regards the sign of charge and that of energy in relation to the direction of propagation in time, as otherwise we may misinterpret ordinary phenomena for potentially forbidden, symmetry violating occurrences.

2.6   The charge conjugation operation

I think that in the context of the preceding analysis it becomes clear that the common assumption that time reversal amounts to simple motion (including rotation) reversal is what prevents a proper understanding of the nature of the charge conjugation symmetry operation. The problem is that if we ignore the dependence of the observed sign of charges on the true direction of propagation in time of the particles carrying them, then this direction of propagation becomes impossible to assert, which explains that the existence of such a degree of freedom has traditionally been ignored altogether. Thus, I believe that the mistake we do when we consider time reversal as it is traditionally defined (even if we can now recognize that this error is not *only* a consequence of using a unidirectional viewpoint) is that we do not consider an evolution according to which the direction of propagation in time of particles is really reversed, but instead consider processes for which a series of events occur forward in time, merely in the reverse order to that in which they would otherwise be observed to occur. But given that non-gravitational charges are not affected by a reversal of the direction of propagation in time of the particles carrying them (which is distinct from the observed direction of their motion) we have a means to determine the direction of propagation in time of particles which therefore becomes a meaningful, well defined concept which must be taken into consideration[3]. It would therefore be incorrect to

---

[3]This conclusion is also justified by the fact that if an observer was 'following' the actual direction of propagation in time of an antiparticle then this antiparticle would appear to have the same charge as its particle counterpart, but then it would be all the

argue that only thermodynamic phenomena allow to distinguish a direction of time (even in the absence of violations of T symmetry), because from a unidirectional time viewpoint the sign of charge is dependent on the direction of time. It is thus simply the fact that the sign of charge itself cannot be characterized in an absolute manner that prevents a direction of time from being singled out as objectively distinct, in the way thermodynamic processes may appear to allow.

Now, what makes the acknowledgement of the existence of a relationship between direction of time and sign of charge unavoidable is the recognized validity of the interpretation of antiparticles as particles propagating backward in time, which allows to identify reversal of time as the very cause of the apparent reversal of charge occurring from the unidirectional time viewpoint. I believe indeed that despite what is often suggested, the interpretation of antiparticles as particles propagating in the opposite direction of time is not merely a helpful analogy with no real significance. Given the absence of a rational motive for rejecting the existence of a fundamental time direction degree of freedom equivalent to the space direction degree of freedom and given the simplification made possible by the discussed interpretation of antimatter in a relativistic context, I think that we must recognize that there definitely exists a relationship between the direction of time and the sign of charge. But it must also be clear that despite what is sometimes proposed there is no equivalence between a reversal of *space* directions and a reversal of the sign of charge (which could imply that antiparticles are merely the enantiomorphic equivalent of their corresponding particles), even if there does occur situations when reversing the space coordinates may appear to counteract asymmetries associated with the sign of charge, because the relationship between space direction and sign of charge is in fact always a consequence of the existence of a relationship between the direction of space intervals and that of time intervals. In any case, if the relationship between time reversal and charge reversal which is suggested by the above mentioned interpretation is considered valid then it would mean that the charge conjugation symmetry operation must actually be understood as itself involving some time reversal.

What I'm proposing therefore is that we should recognize that the charge conjugation symmetry operation C must actually be conceived as a combined

---

other particles in the universe which would appear to have a reversed charge, which is certainly a significant change.

space and time reversal operation that leaves the sign of non-gravitational charges invariant relative to the direction of time in which particles would be propagating following such a reversal. Thus, C must be understood to reverse the time parameter t (associated with the 'position' in time), along with the physical time intervals Δt associated with the propagation of particles, and the energy sign E of those particles (which is reversed as a requirement of action sign invariance). But it must also reverse the space position parameter x, the physical space intervals Δx associated with the propagation of particles, and the momentum p of those particles (which is also reversed as a requirement of action sign invariance). Here again we must recognize that the charge q is actually left unchanged, along with the spin of elementary particles, from a fundamental viewpoint, even by this reversal operation we call charge conjugation. Yet it still makes sense to consider C as a reversal of charge given that, from the viewpoint of unidirectional time, non-gravitational charges would appear to be one of the few physical properties of elementary particles which would actually be reversed by this symmetry operation, while the space and time intervals, along with the energies and the momenta would appear to remain unchanged.

This must happen for the same reasons that justified assuming that momentum and space intervals are reversed by T from a unidirectional time perspective, even though they are left invariant by this symmetry operation from the bidirectional viewpoint. Indeed, upon applying C we are in a situation where all intervals and their conjugate attributes are reversed from a fundamental time-symmetric viewpoint, which means that to satisfy the needs of a unidirectional perspective we must reverse the time-related parameters Δt and E again, but given the relationships that exist between the physical time intervals and the space intervals this means that the space-related parameters Δx and p must also be reversed a second time, just as they were shown to be reversed (once) by T from this unidirectional viewpoint. If the physical time intervals and the energies must be reversed from what they really are (what they have become as a result of applying the operation in the first place) it is therefore due to the fact that from the unidirectional viewpoint we use the wrong direction of time, but given that following time in the wrong direction also implies that the space intervals are followed in the wrong direction (the relational aspect), then this actually means that the space intervals must also be reversed from what they really are (what they have become), along with the momenta. As a result, there appears to be no change to space and time intervals from applying C, even though it

is here defined as a space and time reversal operation. Yet, as charge is not a spacetime related physical property, because it is associated with interactions distinct from gravitation (unlike energy or momentum which can be conceived as the charges determining the metric properties of spacetime), it should be considered that it actually remains unchanged from the fundamental bidirectional viewpoint under a space and time reversal operation such as the properly defined C, which means that it would appear to be reversed, as we would normally expect, from the unidirectional time viewpoint (because time is then followed in the wrong direction).

There is a slight difference, however, between the outcome of a properly defined C operation and the expected outcome of a traditionally defined charge conjugation operation, because the reversal of the space and time position parameters x and t themselves (which now occurs from both the bidirectional and the unidirectional time viewpoint), even if it is without any effect on the sign of the space and time intervals associated with the propagation of particles from a unidirectional viewpoint (given that those intervals must then be reversed a second time), actually implies that angular momentum would appear to be reversed by C (because momentum is indeed unchanged while the position in space is reversed). Thus, despite common expectations, a C-reversed process would also appear to involve reversed angular momentum or spin, which means that contrarily to what is sometimes suggested, the behavior of spin under charge conjugation is not a mere matter of convention and its reversal (apparent from a unidirectional time perspective) must be considered an unavoidable outcome of applying this symmetry operation.

The reversal of spin under C is certainly unexpected according to the traditional approach, but from my perspective it appears natural, given that C involves a reversal of time. It must be clear though that this reversal of spin is only apparent and does not occur at the most fundamental level of description, in accordance with the requirement that an action sign preserving symmetry operation like C should not reverse the sign of action associated with angular momentum. This is to be required even if in general the sign of spin is not uniquely tied to the sign of action associated with energy and momentum, because the only way spin can reverse is when either the position in space or the momentum are independently reversed and an action sign preserving reversal symmetry that reverses momentum would necessarily also reverse spatial position given that it must reverse space intervals (which is not required from the unidirectional time viewpoint relative to which mo-

mentum can be made to vary independently from the sign of space position, even when action is to remain positive).

We are now therefore in the situation where we must recognize that, from a certain viewpoint, charges are reversed by a properly defined time reversal operation T, while spin angular momenta are reversed by a properly defined charge reversal operation C, despite what had traditionally appeared to be required from such discrete symmetry operations. Another distinction of the proposed approach is that handedness is now also reversed by C from whatever viewpoint, because either momentum is reversed and spin is invariant (as from the bidirectional viewpoint), or momentum is invariant and spin is reversed (as from the unidirectional viewpoint), so that there is always a relative change between the direction of spin and that of momentum. The outcome of the proposed charge reversal operation C as it was here redefined would therefore differ from that of a properly defined T operation mainly through the fact that unlike C, T would reverse the momentum and space intervals (from a unidirectional viewpoint), but would not reverse the handedness of particles, just as we would also expect traditionally. Thus, both the P operation and the redefined C operation would alone and from any viewpoint reverse the handedness. In this context the fact that under certain circumstances, such as when the weak interaction is involved, particles of a given handedness seem to be naturally related to antiparticles with opposite handedness could be understood to follow from the fact that the handedness is reversed by a properly defined charge conjugation operation (which still relates particles to antiparticles), so that if there can be invariance under such a symmetry operation then reversing both charge and handedness should not be expected to produce any change. This is an important result which confirms that the suggestion, usually made on the basis of purely phenomenological considerations, that charge conjugation should perhaps involve a reversal of handedness, was in fact justified from a theoretical viewpoint.

2.7 Invariance under combined reversals

I think that I have appropriately justified the inevitability of the above discussed conclusions regarding which parameters should be expected to reverse under the various discrete symmetry operations (in particular when I discussed the requirement of action sign invariance and the constraint of relational definition of the reversal operations), but I must nevertheless mention

how remarkable it is that the described variations of physical parameters under the redefined P, T, and C operations happen to be just such that they *explicitly* require invariance to occur under a combined PTC operation. This happens because all the parameters which are independently reversed by any of the symmetry operations are actually reversed twice when the operations are combined and this regardless of whether we are considering a unidirectional or a bidirectional time viewpoint (a look at tables 2.3 and 2.4 allows to quickly confirm this fact). Either a parameter such as Δt is reversed twice or either it is not reversed a single time by a properly defined PTC and this actually guarantees that there is invariance under a combination of the three discrete symmetry operations, because anything that may be reversed is reversed again and only once. In fact, as I will explain below what we really need is twice a reversal of *all* fundamental space- and time-related parameters (that is both the time-related parameters t, Δt and E, and the space-related parameters x, Δx and p) under a properly defined PTC and this actually occurs when the appropriate bidirectional time viewpoint is considered. Charge and spin on the other hand need not reverse at all from such a viewpoint under a PTC operation as they necessarily transform independently from the action sign preserving discrete symmetry operations and only reverse as a consequence of adopting a unidirectional viewpoint and in such a case they do reverse twice, as required. This is in contrast with the traditional definition of the discrete symmetry operations (described in tables 2.1 and 2.2) according to which some parameters like the space and time coordinates, the charge, and the spin can be reversed a single time only by the combined PTC operation.

We can understand, however, why it is that this combined symmetry operation should be expected to produce invariance even as it is traditionally defined (as required by the CPT theorem). This is possible simply because, according to the traditional conception, while charge would be reversed only once (by C), spin would also be reversed only once (by T), but as one can show, there is a kind of equivalence, at least for fermions, between a reversal of the polarization state associated with spin and a reversal of charge and this is why even under its traditional definition the combined PTC symmetry operation would have to leave physical states invariant (although it would seem to alter the direction of space and time coordinates, which could turn out to be physically significant under particular circumstances). It is also interesting to observe that in the context of my revised definitions of the discrete symmetry operations any two operations applied together is *explicitly*

equivalent to the remaining operation, so that applying PT, for example, is totally equivalent to applying C, which again demonstrates that charge conjugation must really be conceived as a space and time reversal operation and that time reversal must involve a reversal of charge from a certain viewpoint. What those relationships really show is that the discrete symmetry operations as they are now defined are all necessary and together sufficient to provide a complete account of the possible transformations involving a reversal of any of the fundamental properties of matter aside from the sign of action (in fact, as I will explain in section 3.3, charge can also be reversed independently from any space- and time-related attribute, but the states of matter so obtained usually do not interfere with the processes involving ordinary matter and antimatter particles).

In this regard I must also mention that it is not possible to assume that applying either P or T alone but twice should necessarily produce invariance (in the sense that it would leave any system with no discernible change that could be related to unchanged physical parameters) despite the fact that it would appear to leave all parameters unchanged, because such a combined transformation may not leave the quantum phase associated with fermions unchanged given that it would only be equivalent to a rotation in space by 2π radiant (as a single space reversal introduces a π radiant rotation and a single time reversal introduces an equivalent additional π radiant rotation in space) and only twice such a complete rotation would necessarily produce invariance in the presence of fermions. Of course applying P or T alone twice would already be more likely to produce invariance than applying P alone or P combined with T only once, because at least some of the effects of applying P or T once would indeed be neutralized by a second application of the same operation, but the point is that in such a case invariance would not *necessarily* follow. The case of C is different, however, given that this operation involves a reversal of both space and time parameters all at once, which produces an equivalent 2π radiant rotation with only one application (therefore allowing the changes involved to be related to the incomplete transformation of fermion wave functions), so that applying C twice reverses *all* parameters twice and introduces twice a 2π rotation that must leave even the quantum phase of fermions invariant. The C operation as I redefined it is thus unique, because it is the only one of the three relationally distinct discrete symmetry operations that reverses both space- and time-related parameters together and from its alternative definition it can be seen that applying C is actually and explicitly equivalent

to applying a combined PT operation. In this context applying PTC could be considered equivalent to applying PT twice, which clearly shows that the PTC operation involves a reversal of all parameters twice and is also equivalent to two complete rotations, which can only produce invariance.

In fact, any one of the three basic discrete symmetry operations can be considered as equivalent to a combination of the other two, so that T for example would here be equivalent to CP and P would be equivalent to the combined CT. Therefore, applying T twice would be equivalent to applying CP twice, which would amount to reverse both space- and time-related parameters twice (which considered alone would have to produce invariance) and then also reverse space-related parameters twice (the order of application of the discrete symmetry operations in a combined operation has no importance and only the number of times a parameter is reversed is significant). But such a combined operation would not leave fermion wave functions invariant for the same reason that applying P alone twice should not be expected to necessarily leave things invariant. It remains, however, that the fact that some combinations of basic discrete symmetry operations which are not required to necessarily produce invariance do involve twice a reversal of some specific physical parameters, allows one to expect that an invariance which was lost when one of those fundamental operations was applied alone can sometimes be regained by application of such combined operations. This should indeed be expected to occur given that, as I mentioned above, reversing one physical parameter twice, even if it is not guaranteed to leave all processes invariant, still allows the possibility of neutralizing some asymmetries which would occur as a consequence of the reversal of this single parameter.

What must be retained here is that there may be a difference between applying a symmetry operation twice and applying the *outcome* of this operation only once (which would effect no change), even if in certain cases, as when the operation considered is the C symmetry operation, we would necessarily observe no change when the same operation is applied twice. This particularity of the C operation is merely a consequence of the fact that it reverses more individual parameters all at once so that applying it in combination with itself actually allows to leave no parameter unchanged relative to which an asymmetry could be properly defined. It must be understood, however, that despite their equivalence with combinations of distinct operations, the three basic operations defined above are all essential to a description of the allowed discrete transformations of physical parameters and none is more

fundamental than any other. Indeed, two operations are distinct from a relational viewpoint, when one of them reverses one category of parameter, say space, relative to the other category, say time, while the other reverses another category of parameter, say time, relative to the previous one, say space, and each one of those operations is relationally distinct from yet another one that reverses both categories of parameters together and which constitutes the necessary complement to the other two operations.

2.8   The significance of classical equations

We can now return to the problem of understanding how it is possible for the momentum p to be left unchanged by a properly defined time reversal operation T which from the most fundamental viewpoint must be assumed to reverse time intervals dt, but to leave space intervals dx unchanged. A problem would in effect appear to arise from the fact that according to the classical equation that defines the momentum of a particle with mass m we should have $p = m\,dx/dt$, which would clearly imply that if dt is reversed or negative while dx is invariant or positive then p should be negative, which is contrary to my proposal that both space intervals and momentum are unaffected by a reversal of time. But I would like to suggest that this contradiction is only apparent and a result of the fact that the classical equation for momentum is actually valid only from a unidirectional time viewpoint, because it was originally introduced under the implicit assumption that physical properties are always measured in the conventional forward direction of time.

Indeed, what the classical equation is telling us is merely that from the unidirectional viewpoint of an observer always following events in the unique direction of time associated with entropy increase and providing an account of physical quantities like momentum and space intervals in relation to that unique direction of time, relative to which time intervals dt are in effect positive definite, independently from the true direction of propagation in time of the particles involved, some quantities like dx which we might assume not to be reversed by T are actually observed to be reversed while dt itself is kept unchanged. Thus, if we use the viewpoint relative to which we are allowed to assume that the above equation is valid then dt would actually remain positive definite despite the reversal of time, while dx would have to be assumed reversed (for reasons I have already explained), which means that momen-

tum would also be reversed according to this action sign preserving classical equation, which agrees with the definitions I provided for the unidirectional time viewpoint and which is certainly appropriate given that particles submitted to such a time reversal operation must have unchanged momentum in the apparent (but false) direction of their motion, which is satisfied when both the momenta and the physical space intervals are together reversed.

There is no contradiction here, despite the fact that we must assume that the true signs of conjugate physical parameters such as the space intervals and the momenta are together invariant under a reversal of time from the alternative time-symmetric viewpoint (according to which the sign of time intervals is itself reversed), because in such a case the classical equation no longer applies, simply because as a traditional formula it *never* really applied to such situations. The classical relation between momentum and the space and time intervals was deduced on the basis of the validity of a thermodynamic viewpoint of time and therefore does not apply in a context where time intervals are allowed to change sign. The classical equations are logical deductions dependent on a certain viewpoint of time which must be considered inappropriate at the most fundamental level of description. In other words, it is not the validity of the classical equations in a limited context which implies that the assumptions made from a time-symmetric viewpoint (concerning the sign of physical quantities) are contrary to experimental evidence, but really the limited value of the classical equations which imply that the assumptions associated with a unidirectional viewpoint are not generally valid. We must recognize that the assumptions used in the more appropriate time-symmetric context regarding the variations of space- and time-related quantities under a reversal of time are not just theoretically well motivated, but that under the right interpretation they are fully supported by observations, while the variations deduced from a unidirectional time viewpoint are explainable merely in the context where they are assumed to derive from the more fundamental bidirectional description.

It must be clear that in this context we would also be unjustified to make use of the classical formula for angular momentum $\boldsymbol{L}$, to which the spin of elementary particles is related, to decide what would happen to spin from a fundamental viewpoint under a reversal of time effected by a properly defined T or C operation. Indeed, the classical formula defines the angular momentum $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$ in terms of the position vector $\boldsymbol{r}$ and the momentum $\boldsymbol{p} = m\,(dx/dt)\,\boldsymbol{i}$ and if we assume a reversal of time intervals dt to follow from both a T and a C reversal operation then according to this equation it would

seem that L should reverse under both types of time reversal, because either dt reverses alone (as under a properly defined T) or it reverses along with r and dx (as under a properly defined C). But, as I already mentioned, and for reasons I have previously discussed, it would be incorrect to assume that angular momentum reverses under either T or C from the bidirectional time viewpoint relative to which dt does in effect reverse. Yet there is no problem here, because the classical formula is only right when we consider things from the unidirectional viewpoint according to which dt is positive definite, but under such conditions either dx and p reverse together with unchanged r (as occurs when T is applied), or else dx and p are unchanged and r is reversed (as occurs when C is applied and only space positions are reversed) so that in both cases spin angular momentum should actually reverse. Again it must be emphasized that the incompatibility of the classical equation for angular momentum with the definition of time reversal as it occurs from a fundamental bidirectional viewpoint must not be considered to imply that the proposed fundamental definition is inapplicable, because all that it means is that the equation itself is of limited scope, having been developed in the context of a unidirectional perception of the evolution of physical systems, when it had not yet even been realized that there exists a fundamental degree of freedom associated with the direction of propagation in time.

2.9   Reversal of action

The clarification of the situation which was achieved in the preceding sections regarding the interdependence of fundamental physical properties as they vary under application of any of the three essential discrete symmetry operations has allowed to establish that that none of the traditionally considered discrete symmetry operations engenders a reversal of the sign of action. This is of course a consequence of the fact that regardless of the viewpoint we adopt, those symmetry operations always reverse the sign of energy in combination with the sign of time intervals associated with the propagation of particles, just as they always reverse the direction of momentum in combination with the direction of space intervals. Thus, the T operation in particular, despite the ambiguity of its traditional definition, cannot be assumed to reverse the action, because while it reverses the time position parameter and leaves the sign of energy unchanged from the unidirectional time viewpoint, it is also implicitly assumed to preserve the sign of time in-

tervals associated with the propagation of elementary particles. The role of inverting the sign of action must therefore be attributed to some symmetry operations distinct from all of those which are usually considered.

I have come to understand that there is not a unique single operation relating positive and negative action states, but that there are basically four different ways by which action can be reversed, which give rise to four different action sign reversing symmetry operations, whose four different outcomes are each related to phenomenologically distinct states of negative action matter. If any one of those operations is applied independently from the others, it may not necessarily produce invariance. I will collectively denote those operations by the letter M to emphasize the fact that they constitute a different category of reversal transformations which are unlike those already studied. The states produced by those four distinct operations can be transformed into one another by individually applying each of the three action sign preserving symmetry operations P, T, and C and therefore I will denote the various action sign reversing operations by applying the appropriate indexes corresponding to the operations which relate the states they generate to the state which is produced by one of those action sign reversing operations chosen arbitrarily as the basic operation, which will itself be denoted M_I. The four discrete symmetry operations so defined are thus the M_I, M_P, M_T, and M_C operations displayed in table 2.5. It must be clear, however, that the choice of which action sign reversing transformation must be associated with the basic operation M_I is completely arbitrary and we could, for example, have defined the operation originally denoted M_C to be the basic operation, which we would instead denote M_I' and we would then obtain the states produced by the other three operations by applying P, T, and C to the state generated by M_I'. That way it would appear that it is the redefined M_C' which would be equivalent to the original M_I, while M_P' would be equivalent to M_T, and of course M_T' would be equivalent to M_P and therefore we see that attribution of the indexes is purely a matter of convention. The letter M was chosen to denote action reversal because the operations it represents would actually alter the gravitational properties of the matter submitted to such reversals and mass (which is usually denoted m) is the property that was traditionally associated with the gravitational interaction.

From the tables it is possible to see that there are two different ways by which a given type of fundamental physical parameter, either space- or time-related, can be reversed in such a way that the sign of action is reversed. We can either assume a reversal of the signs of momenta and energies relative to

Bidir.	t	Δt	E	x	Δx	p	q	s	h
$M_I = M'_C$	t	Δt	$-E$	x	Δx	$-p$	q	$-s$	h
$M_P = M'_T$	t	Δt	$-E$	$-x$	$-\Delta x$	p	q	$-s$	$-h$
$M_T = M'_P$	$-t$	$-\Delta t$	E	x	Δx	$-p$	q	$-s$	h
$M_C = M'_I$	$-t$	$-\Delta t$	E	$-x$	$-\Delta x$	p	q	$-s$	$-h$

Table 2.5: Variations of physical parameters under the four relationally distinct action sign reversing symmetry operations as described from the bidirectional time viewpoint. Here I chose the basic action reversal operation M_I to be that which reverses energy E independently from time intervals Δt, and momentum p independently from space intervals Δx. Under an equivalent definition it would be the time intervals Δt and the space intervals Δx which would be reversed by the basic action reversal operation M'_I while the energy E and the momentum p would be kept invariant.

unchanged space and time intervals or we can assume a reversal of the space and time intervals associated with the propagation of particles that would occur while keeping the signs of momenta and energies invariant. But given that those two different kinds of reversal can be applied differently to space- and time-related parameters (you can apply one kind of reversal to space and the other to time or vice versa as long as you do apply any one type of reversal to each type of parameter), it means that there are four different kinds of operations in all which can reverse the sign of action. From those definitions it is clear that what the M_I, M_P, M_T, and M_C operations really involve is the reversal of an additional degree of freedom relationally distinct from those already affected by the P, T, and C operations, because indeed even the state obtained by applying the basic M_I operation actually involves a reversal of action, which means that all possible states related by application of P, T, and C, including the original state obtained by application of the identity operation I have their counterpart as M-reversed states and under such conditions we can only conclude that we are actually dealing with a transformation that applies to a distinct property of matter. The illustration of the effects of the various action sign reversing operations depicted in figure 2.3 allows to clearly identify this degree of freedom as the relative orientation of momentum p compared to space intervals Δx or equivalently that of energy E compared to time intervals Δt, which for negative action states is the

opposite of what it is for positive action states.

The C, P, and T operations, therefore, do not together operate a reversal of all fundamental physical parameters, because they merely reverse all parameters while leaving the sign of action invariant. The four action sign reversing symmetry operations proposed here are then the additional operations which are required to complete the set of discrete space- and time-related symmetry operations, because they perform the only remaining possible changes that the traditional operations do not produce, by actually reversing the sign of momentum and energy relative to the direction of space and time intervals. From that viewpoint it appears that even though they are usually ignored the M_I, M_P, M_T, and M_C operations cannot in fact be avoided. The fact that there are actually four distinct operations that can perform a reversal of action on the other hand simply means that it is not possible to associate a unique state of momentum or energy, or of propagation in either space or time, to negative action matter and that all the different action sign preserving variations of the direction of fundamental physical parameters which can apply to positive action matter would also apply to negative action matter. We can thus actually expect that there would, for example, be a charge conjugation symmetry operation C applying independently to negative action matter, which would therefore have its own antimatter particles distinct from ordinary antiparticles.

In this context it appears that the distinction that exists between matter and antimatter must be attributed essentially to the true direction of propagation in time of particles, independently from their sign of action. An antiparticle is therefore always just a particle which reversed its energy while changing its direction of propagation in time, which is not very different from the situation of a particle which reverses its momentum by changing its direction of motion in space. Indeed, by reversing its momentum when it changes its direction of propagation in *space* a particle is allowed to keep the sign of its momentum relative to the direction of its motion unchanged, so that its action sign is also unchanged, just like a positron retains the sign of action of the electron with which it annihilates, because the electron reverses its energy when it starts propagating backward in time (which is viewed as the annihilation process forward in time). But a negative action particle would be clearly distinct in this respect as a consequence of the fact that it would have not only negative energy carried forward in time (or positive energy carried backward in time, which is equivalent from a unidirectional time viewpoint when the sign of charge can be ignored), but also negative momentum in the
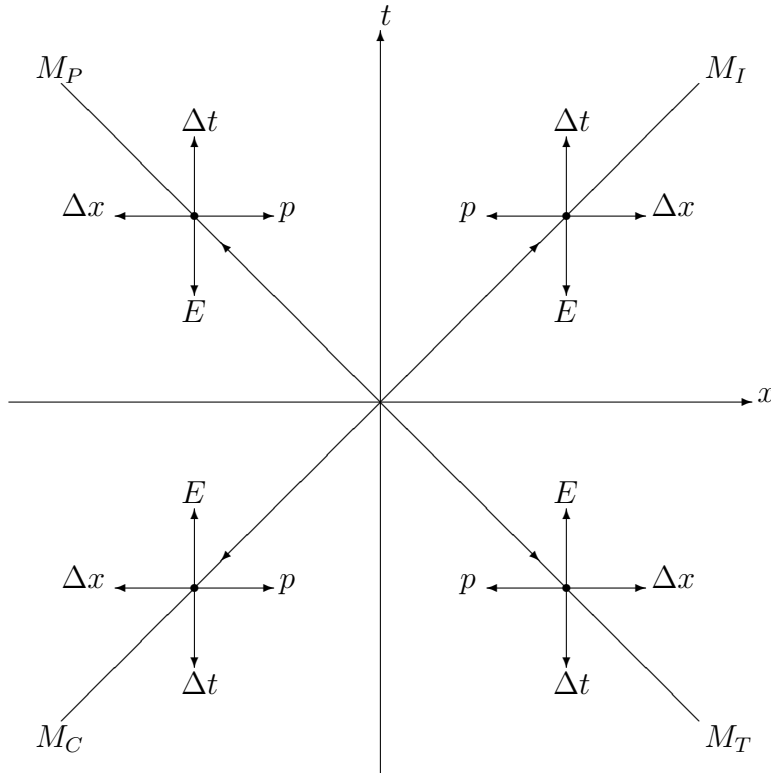
Figure 2.3: Four different outcomes of applying each of the relationally distinct action reversal symmetry operations as described from the bidirectional time viewpoint. Here we notice that the orientation of the vectors which correspond to the signs of space and time intervals is always opposite that corresponding to the signs of momentum and energy, as we should expect to observe when action is indeed negative. If we were to consider a unidirectional time viewpoint we would have to reverse all space and time intervals and all momentum and energy signs for the processes obtained by application of both the M_T and M_C operations, which means that all four operations would give rise to the propagation of negative energies forward in time.

observed direction of its propagation in space (the momentum would point in the direction opposite the observed velocity of the particle), unlike any ordinary matter particle (including antiparticles). It must be clear, however, that according to the proposed definition of action sign reversing symmetry operations which is described in table 2.5, non-gravitational charges are assumed to be unaffected be a reversal of action, just as they were left invariant by the action sign preserving reversal operations. Only the practical necessity of a forward in time viewpoint would for negative action matter also imply that charges appear to be reversed when a process is submitted to an action sign preserving reversal of time.

Another particularity of the operations of action reversal defined above is that spin is deduced to be reversed under all such relationally distinct operations when their effects are considered from the bidirectional time viewpoint. This is certainly just as appropriate as is the invariance of spin observed for all action sign preserving symmetry operations, because as I previously mentioned spin has the dimension of an action and should therefore vary in correspondence with the sign of action associated with momentum and energy from a fundamental viewpoint. The constraint on the variation of the direction of spin is actually the same constraint that requires that either both space- and time-related parameters are such as characterizing a positive action state, or else that they are both such as characterizing a negative action state and that it should not be possible for one single particle to propagate, say, positive momentum in the direction of its motion in space and at the same time propagate negative energy forward in time. This is a simple matter of consistency, because a physical system cannot have at once both the gravitational properties associated with positive action matter and those associated with negative action matter if, as I suggested in the previous chapter, the attractive or repulsive nature of the gravitational interaction between two particles actually depends on the difference or identity of their action signs. This does not mean, however, that spin cannot vary independently from the sign of action associated with energy and momentum, but merely that while it cannot reverse as a consequence of applying an action sign preserving discrete symmetry operation, it also *must* reverse as a consequence of applying a reversal of action.

It may also be noted that just as is the case for the action sign preserving discrete symmetry operations, some combinations of two of the four operations describing a reversal of action are equivalent to a combination of the other two operations (in the case of the action sign preserving operations

one operation, which is that of charge conjugation C, was equivalent to the other two, but in fact this single operation was implicitly combined with the invariant operation I which effected no additional change and thus could be ignored). Here a combination of M_I and M_C or a $M_I M_C$ operation would be equivalent to a combination of M_P and M_T and this is what allows a combination of all action sign reversing symmetry operations (or a $M_I M_P M_T M_C$ operation) to necessarily produce invariance, given that all relevant parameters are actually reversed twice by such a combined operation. In fact, it turns out that combining any of M_P, M_T, or M_C with M_I produces an operation equivalent to the above defined P, T, or C respectively (while a combination of M_I with itself produces an operation equivalent to the identity operation I) so that a combination of the other two remaining action sign reversing operations would also be equivalent to those action sign preserving operations. For example, the combined $M_P M_T$ operation is mathematically equivalent to a C operation because it reverses both space- and time-related parameters once and reverses the action twice, which is equivalent to leave action unchanged.

One must understand, however, that even though applying any one action sign reversing operation twice would be equivalent to applying the identity operation I, such a combined operation would not necessarily produce invariance and this for the same reason that applying P or T twice would not necessarily leave everything invariant despite the fact that it would also appear to be equivalent to applying the I operation, which effects no change. This is, again, because applying an operation that does not reverse all physical parameters twice, even if it may appear to return a system to its original state, may still produce a change which can be characterized in a relational way, because some parameters would be reversed relative to other parameters which remain unaffected by the transformation and this may not leave the processes involved invariant. Still regarding the conditions for necessary invariance, it should be clear that simply combining a PTC operation with the basic M_I or any other action sign reversal operation as a way to try to regain invariance which may be lost upon reversing the action (in the way we would apply T to a CP violating process) cannot be expected to produce invariance given that the action sign degree of freedom would then be reversed only once. Thus, a violation of any of the M symmetries would not imply that there must be a violation of PTC symmetry, as we may understand to be independently required on the basis of the fact that invariance under PTC alone must itself be considered unavoidable. The appropriate generalization

of the PTC (or really $IPTC$) symmetry must then be recognized to be the $M_I M_P M_T M_C$ symmetry which combines all the relationally distinct action reversal symmetry operations and which must therefore (because there would remain no unchanged physical parameter relative to which a change could be determined) be equivalent to no change at all. Indeed, as indicated in table 2.5, a physical parameter may either not be reversed by any of the action reversal operations or else be reversed by two or all of those symmetry operations, which explicitly guarantees invariance under a combination of the four operations.

Now, in order to avoid confusion, it is important to understand that the action sign reversal symmetry operations must be considered as operations distinct from one another that apply to an identical state, rather than as an identical operation that applies to different states. In such a context it transpires that the fact that the M_I, M_P, M_T, and M_C operations are related to one another through application of the various action sign preserving symmetry operations merely shows that the states obtained by applying the four action sign reversing operations are themselves related to one another through the same action sign preserving operations that transform unchanged action sign states into one another. Thus, despite the fact that all of the action sign reversing symmetry operations are equivalent to a combination of some arbitrarily chosen action sign reversing operation with one of the four action sign preserving operations, it would not be appropriate to assume that invariance can be obtained by applying each of the action sign preserving symmetry operations along with a single particular action sign reversing operation to obtain a $(M_I I)(M_I P)(M_I T)(M_I C)$ operation (which would not necessarily produce invariance despite the fact that $IPTC$ must leave everything invariant).

What must be clear is that no action sign reversing symmetry operation can be identified as *the* action reversal operation and under such circumstances it is not possible to avoid having to consider the many operations as distinct from one another despite the fact that all such operations can be obtained by combining in turn each of the action sign preserving symmetry operations with just one single action reversal operation. In this context it is important to realize that action reversal symmetry can be violated to different degrees when one transforms a state of positive energy matter into the different states of negative energy matter which are related to one another by the redefined action sign preserving reversal operations P, T, and C, because each of those states is related to a corresponding state of positive energy

matter by a specific action sign reversing symmetry operation and these operations do not necessarily produce invariance when applied separately. Thus, the P, T, and C operations can be violated to different degrees by negative energy matter (compared to how they are violated by positive energy matter) when applied independently from one another and this precisely because M_I, M_P, M_T, and M_C can themselves be violated to different degrees in comparison with one another, so that they relate the different asymmetric states of positive energy matter to corresponding states of negative energy matter which can be asymmetric in different ways relative to one another. The only requirement is that the different states of negative energy matter which are related to the different states of positive energy matter by the various action sign reversal symmetry operations be subject to the same invariance under a combined PTC transformation as are states of positive energy matter, even if P, T, and C are violated to different degrees by negative energy matter in comparison with the violations occurring for positive energy matter. The four action reversal symmetry operations, therefore, simply allow to relate all the positive energy states which are transformed into one another by the action sign preserving symmetry operations to all the negative energy states which are transformed into one another by similar operations. Thus, despite the existence of four distinct action sign reversal symmetry operations, action reversal must really be conceived as transforming one single degree of freedom and this means that I'm justified in referring to the action reversal operations collectively as the M symmetry.

In any case it appears that the commonly met remark to the effect that gravitation is invariant under a reversal of time must be nuanced. What I mean is that while it is certainly true that there would be no change to the attractive or repulsive nature of the gravitational interaction if time was locally reversed for some physical system by a time reversal operation such as T, we should certainly expect a reversal of time independent from the sign of energy, such as that produced by an M_T operation, to exert a change on the nature of the interaction of the affected system with the rest of the universe. Indeed, such a transformation would reverse the sign of action and as I previously explained the repulsive or attractive nature of the gravitational force between two bodies depends on the relative value of their action signs (because gravitation is always attractive only for particles with the same sign of action). But, even if we consider a reversal of time as produced by an action sign reversing operation like M_T to apply to the whole universe (in which case we would have to use negative energy matter in place of positive energy

matter when testing for invariance), the preceding discussion made clear that we should not necessarily expect to observe phenomena which would be entirely identical with those of the original universe, because M_T applied alone could be violated, just as any operation which is not reversing all physical parameters twice. This would also be true of M_P for example, because just as the change in the sign of time intervals produced by an M_T operation can be related to an unchanged sign of energy, so the change in the direction of space intervals produced by an M_P operation can be related to an unchanged direction of momentum.

Yet the fact is that there *could* in effect be invariance under a reversal of time that does not preserve the sign of action if the operation is applied to all particles in the universe, because in such a case the difference or the identity of the signs of action of the various particles would not be affected and this is the only aspect that would be significant from a gravitational viewpoint. But this invariance would apply only to the extent that there is in effect no violation of symmetry under exchange of positive and negative action states. It is important to mention, however, that even if one might be tempted to conclude, based on a certain interpretation of the generalized gravitational field equations which were proposed in section 1.15, that the minute imbalance which is responsible for the observed small, but non-vanishing positive value of the cosmological constant arises from such a violation of M symmetry, this would not be a valid conclusion, because, as I will explain in section 3.2, this imbalance rather develops as a consequence of the fact that the rates of expansion of space experienced by observers of opposite energy signs are allowed to differ as time goes, even if they were initially the same and this can occur even in the absence of a violation of M symmetry. Also, as I have explained in the preceding sections of this chapter, simply reversing the direction of motion of particles cannot be considered to consist in a true time reversal operation in any meaningful way, so that assuming that such a transformation would leave all processes unaffected, even when gravitation is involved, could not be understood to mean that gravitation is invariant under time reversal.

2.10   The problem of matter-antimatter asymmetry

It has been suggested more than once that the violations of CP symmetry which have been observed in certain experiments and which are believed to imply a violation of time reversal symmetry T could perhaps be the cause of the observed thermodynamic time asymmetry in our universe. It is usually recognized, however, that the weakness of the violation of T symmetry that is involved would prevent it from being responsible for such an extreme difference between past and future evolution as that which gives rise to the thermodynamic arrow of time. A less common proposal is that it might be the thermodynamic time asymmetry itself which is giving rise to the violation of T symmetry. But this intrusion of macroscopic physics into the affairs of microscopic quantum processes is usually not believed to be a likely possibility, at least by those who do not expect a complete overthrow of conventional particle physics. In fact, I think that what really justifies this attitude is the recognition that what currently remains unexplained is the thermodynamic arrow of time, while any fundamental time asymmetry observable at the elementary particle level could be accommodated by the same rules that make violations of parity possible. Indeed, I have already explained why we do not need to appeal to thermodynamic time asymmetry to legitimize the violation of T symmetry, given that it is allowed to occur as a relationally defined asymmetry (it does not need to be defined relative to the direction of thermodynamic time as it is already defined in relation to other fundamental direction parameters). What's more, the solution I will propose in chapter 3 to the problem of the origin of thermodynamic time asymmetry appears to be incompatible with the hypothesis that the violations of CP symmetry which are observed to occur in certain processes involving elementary particles could be a consequence of the constraint which is actually giving rise to the existence of the thermodynamic arrow of time.

Now, even if thermodynamic time asymmetry is probably not the cause of violations of T symmetry, the direction of time singled out by T violations can be related to the macroscopic arrow of time and this might allow one to conclude that our universe is characterized by a phenomenologically apparent fundamental lopsidedness. Given that the time reversal symmetry operation can now be understood to involve a transformation of matter into antimatter, the question of whether there actually exists such a preferred direction in time

would be equivalent to ask if there really is an absolutely definable asymmetry between matter and antimatter in our universe. However, when I examined this question in the light of the more appropriate conception of antimatter which arose from the developments featuring in the preceding sections I found out that there is after all no absolutely definable lopsidedness if we recognize the validity of a certain hypothesis concerning the continuity of the flow of time along a particle's world-line.

This hypothesis is that which I had at some point contemplated as potentially offering the required constraint that would prevent transitions in which the direction of propagation in time of a particle reverses without being accompanied by a reversal of the energy of the particle (thereby giving rise to processes of creation and annihilation of pairs of opposite action particles out of nothing). I mentioned in the discussion of this problem that appeared in the previous chapter that this condition of continuity must in effect prevent certain changes from occurring on a particle world-line, even though all by itself the limitation involved is not restrictive enough to prevent a reversal of energy independent from the direction of propagation in time (a reversal of action). I'm now allowed to assert that what such a condition of continuity requires, in effect, is merely that there needs to be a continuous flow of the fundamental time direction parameter associated with the sign of physical time intervals along an elementary particle world-line in spacetime. This restriction becomes relevant in the context where it is recognized that there does exist a fundamental time direction degree of freedom distinct from the observed direction of motion of elementary particles.

Compliance with such a continuity requirement would imply that any particle-antiparticle annihilation process, whether it involves particles with the same action sign, or particles with opposite action signs can only occur as the kind of events during which a particle bifurcates in spacetime to start propagating in the opposite direction of time and not as a chance encounter of two opposite-charge particles propagating in the same direction of time. This requirement would then also impose that events cannot occur which would appear to involve a particle turning into its antiparticle by releasing twice its charge without ceasing to exist from the unidirectional time viewpoint, because such processes would imply that the continuous path of a particle in spacetime (the arrow along a particle world-line) could come to an end as a consequence of a particle by chance meeting its backward propagating antiparticle from the future.

Yet we have no choice but to assume that *ordinary* antiparticles (those

that routinely take part in interactions involving ordinary matter) are indeed backward in time propagating particles (and not particles propagating opposite charges forward in time), because, as I mentioned in the discussion concerning the time-direction degree of freedom appearing in section 1.2, if we are to view any transformation along a particle world-line as a continuous process then given that the annihilation of an ordinary particle with an ordinary antiparticle must be allowed to occur with the same probability for all such pairs and cannot only take place for those pairs where the two particles would happen to be those propagating in opposite directions of time, then ordinary anti-particles must always be considered to propagate in the direction of time opposite that in which the corresponding particles are propagating. Thus, even if some of the electrons that propagate in a particular direction of time could have negative charge, while others would have positive charge, we must consider as empirically forbidden for particles and antiparticles with such opposite *bidirectional charges* (the invariant measures of charge which are not affected by a conventional reversal of the direction of propagation in time of elementary particles) to transform into, or to interact with one another and in the context where a condition regarding the continuity of the flow of time is required to apply along a particle world-line this means that no particle can turn into an antiparticle without actually reversing its direction of propagation in time at the instant where the transformation event takes place (therefore describing an ordinary particle-antiparticle annihilation process from the unidirectional time viewpoint) even if charge could perhaps be conserved when a particle would turn into its antiparticle without bifurcating in time (through the emission of a compensating amount of charge carried by interaction bosons).

What I would like to suggest, therefore, is that we must consider as a necessary rule rather than as a convenient assumption that the arrow associated with the direction of propagation in time of a matter particle (from a *bidirectional* time viewpoint) can never reverse along a continuous particle world-line in spacetime. This requirement can be formally expressed using the following definition.

> **Condition of continuity of the flow of time**: There must always be continuity in the direction of propagation along a particle world-line for elementary fermions, so that a particle cannot turn into an antiparticle (and vice versa) without changing its direction of propagation in time in such a way as to preserve the

continuous flow of the fundamental time direction parameter.

If this assumption is valid then from the unidirectional time viewpoint a particle cannot appear to continue propagating forward in time after changing into its antiparticle (just as from the same viewpoint an antiparticle could not have kept propagating backward in time *before* an event at which it transformed into its particle counterpart). Therefore, if a particle continues to propagate forward in time then it must actually retain the sign of its charge, because if it does not then either the condition of continuity of the flow of time would be explicitly violated or we would have to assume that a particle propagating a positive charge forward in time could sometimes transform into a similar particle propagating an opposite charge in the same forward direction of time while, as I mentioned above, this must be considered to be empirically ruled out in the context where the condition of continuity of the flow of time must apply (because annihilation processes involving ordinary matter and antimatter do not only occur for a subset of particle-antiparticle pairs). In section 3.3 I will explain what justifies the validity of the empirical rule that particles propagating a given charge forward or backward in time cannot transform into, or interact with similar particles propagating an opposite bidirectional charge forward or backward in time.

It must be clear, however, that this limitation is not currently recognized as a requirement of elementary particle theories, even though no process that violates this rule has ever been observed. In fact, some unconfirmed grand unification theories actually predict the existence of processes which would violate this continuity condition, but in my opinion this is probably reason enough to doubt their validity given the awkwardness of the kind of evolution they would describe in the context of the best interpretation we have for the nature and the origin of antiparticles. It must be clear, in any case, that even when the proposed constraint applies, the charge of a particle (not necessarily the electric charge) can still change on a continuous world-line (as when a blue quark turns into a red quark, or a neutrino turns into an electron), because all that is required is that a particle does not change into its own antiparticle on such a continuous path (the charges cannot reverse), particularly in the case of fermions, so that if a particle was propagating forward in time, it can still be assumed to propagate in the same direction after the transformation has occurred.

To summarize, if the transformation of a particle into an antiparticle (or vice versa) could occur forward in time on a continuous world-line then there

is no way such a hypothetical transformation could be described as an actual change in the particle's properties at the point in time where the transformation occurs, because the phenomenon could only be appropriately described as the encounter of two independent particles approaching the same event from opposite directions in time and nevertheless meeting at a very specific point in space and this is precisely why the condition of continuity of the flow of time could not be satisfied in such a case. But if we allow for such discontinuity we would then require unlikely coincidences (involving the coordination of distinct forward and backward particle propagation processes) to produce the required meeting of world-lines without any local causality being responsible for this otherwise improbable coordination. Such events would not even be explainable in the way we could explain the chance meeting of two distinct particles at a point in space which would need to occur in the case of a traditional particle-antiparticle annihilation process inappropriately described as the encounter of two opposite-charge particles both propagating forward in time. I believe that those difficulties alone provide enough justification for assuming that a condition regarding the continuity of the flow of time must be imposed under all circumstances.

Now, if the kind of processes just described cannot occur, then it becomes possible to predict that there should be as many forward in time propagating particles as backward in time propagating particles, so that there should be no fundamental lopsidedness involving the direction of time in our universe. Indeed, if we impose as a condition that there must be continuity of the direction of the flow of time along an elementary particle world-line in spacetime, then any forward in time propagating particle present at a given moment must be accompanied by a corresponding backward in time propagating particle, because no forward in time propagating particle can be created without its backward in time propagating counterpart also being created in the process. Of course this conclusion would be valid only under the assumption that all matter particles present in the universe must in effect be created from nothing at the Big Bang. But even if this assumption may not appear appropriate in the context where time would somehow extend past the Big Bang following a hypothetical Big Bounce (given that in such a case matter could perhaps already be present in the universe that would not need to be created), in sections 3.5 and 3.9 I will explain that creation out of nothing is actually an unavoidable requirement that must apply regardless of whether time extends past the initial singularity or not.

But even if highly suitable, the conclusion that there must be as many

particles propagating forward in time as there are particles propagating backward in time may at first seem problematic, because there is in effect more matter than antimatter in our universe. This observation probably explains why it had never been considered that a condition of continuity of the flow of time applying to elementary particle world-lines may actually impose that the number of forward in time propagating particles be equal that of backward in time propagating particles. I would like to suggest, however, that in the context where the existence of negative energy matter is recognized to be unavoidable, the absence of antimatter in our universe would not rule out the validity of the above discussed conclusion, because we are allowed to assume that the number of backward in time propagating particles could be larger than that of forward propagating particles for negative action matter and if that is actually the case then the condition of continuity of the flow of time, which requires equal numbers of forward and backward in time propagating particles could still be satisfied despite the observed asymmetry between positive action matter and antimatter. The truth would then simply be that the matter-antimatter asymmetry is reversed for negative energy matter (despite the fact that negative energy observers would likely refer to particles propagating forward in time as their own antimatter if those particles are less abundant than backward propagating particles) and that there is actually the same number of otherwise identical positively and negatively charged particles (as observed from the viewpoint of thermodynamic time) when we appropriately take into account the contribution of the unseen negative energy matter. This would in effect be allowed in the context where (as I previously explained) it seems possible for particles of opposite action signs to be permanently created together under the conditions which prevailed during the Big Bang.

On the basis of the preceding arguments it appears necessary to assume that negative action matter is mostly composed of protons and electrons with charges opposite (from the forward in time viewpoint) that of our most abundant protons and electrons, a conclusion which is particularly appropriate in the context where any reversal of the sign of action is assumed to leave charge invariant, so that the opposite directions of propagation in time of the most abundant forms of positive and negative action matter should alone determine any difference in the sign of their charges that would be apparent from the unidirectional time viewpoint. What's interesting is that this regained equilibrium between matter and antimatter would have to be observed regardless of the exact nature of the phenomenon which is responsi-

ble for the violation of T symmetry that gives rise to the imbalance affecting positive action matter when it is considered independently from negative action matter. Thus, if the above defined condition of continuity of the flow of time along an elementary particle world-line is valid, the number of ordinary matter particles may still be allowed to change independently from that of ordinary antiparticles, but only when there is an opposite variation in the relative number of negative action matter particles over negative action antimatter particles such that the total number of matter particles of *all* action signs remains rigorously equal to the total number of properly defined antimatter particles of all action signs.

What I'm suggesting in effect is that this compensation of the observed matter-antimatter asymmetry made possible by the presence of negative action matter is not just a mere possibility, but that the requirement identified above actually implies that there must necessarily be an equal number of forward and backward in time propagating charges when all possible forms of matter are considered together. The direction of entropy growth in our universe would thus correspond to the direction of propagation in time of the most abundant form of positive action matter, but also to the direction of propagation in time of the less abundant form of negative action matter and this contributes to somewhat restore the required symmetry that is lost as a consequence of the existence of a thermodynamic arrow of time. The apparent asymmetry between matter and antimatter would merely be a consequence of the fact that the presence of appropriately conceived negative action matter is not taken into consideration by traditional models. The plausibility of the identified requirement concerning the continuity of the flow of time along the world-lines of elementary particles is therefore what allows me to conclude that the matter-antimatter asymmetry characterizing our universe cannot be used to identify a fundamental lopsidedness in time (assuming that there is a correspondence of the thermodynamic arrows of time, independent of the sign of time intervals, for positive and negative action matter). But, given that I have argued (based on independent motives) that it is not possible for absolutely (non-relationally) defined space and time directions to occur at a fundamental level, then we may consider that the solution to the issue discussed here is a confirmation of the validity of the hypothesis involved.

It must be understood, however, that if the universe is required to be invariant under PTC, as is unavoidable under the above proposed alternative definition of those discrete symmetry operations, then any asymmetry asso-

ciated with the sign of charge would have to be compensated by asymmetries associated with other physical parameters of matter with the same sign of action, because none of P, T, or C involve a reversal of action. Invariance under PTC cannot in this context be invoked as possibly requiring a compensation of the matter-antimatter asymmetry affecting positive action matter by some asymmetry involving negative action matter, because invariance under PTC is preserved independently from invariance under $M_I M_P M_T M_C$ and therefore only the above described condition of continuity of the flow of time actually requires that there is a compensation between the lopsidedness of positive energy matter and that of negative energy matter. It should be noted as well that if the above conclusion is valid then we should expect that there would not only be an equal number of forward and backward propagating matter particles of any type, but also that there would be an equal number of positive and negative action particles. Given that the sign of action is the significant parameter when the gravitational interaction is concerned, this can be considered appropriate despite the fact that the discussed constraint would imply that there are actually more positive energy particles than there are negative energy particles propagating in any direction of time (because there would then be more positive action particles propagating positive energy forward in time, but also more negative action particles propagating positive energies backward in time). It would then remain to establish if the prediction that there should be an equal number of positive and negative action matter particles in our universe is viable from an observational viewpoint. I will return to this important question in chapter 3.

One less obvious consequence which would emerge if the above proposed solution to the problem of matter-antimatter asymmetry is valid is that there should then *necessarily* exist conditions in our universe under which positive and negative action states could transform into each other when the appropriate reversal of the direction of propagation in time is involved. This is indeed a consequence of requiring a continuity of the flow of time along the world-lines of elementary particles, because given the observed imbalance between the number of ordinary positive action particles and that of antiparticles it must be assumed that some of the bifurcation points in time involve pairs of opposite action particles. This may appear to contradict the previously discussed conclusion that processes of pair creation involving opposite action particles cannot occur as permanent outcomes under ordinary conditions (because the particles so created would immediately annihilate back to nothing), but the only conclusion we can draw from the above analysis is that it must

actually be possible, when the scale considered is that of quantum gravitation and space is expanding sufficiently rapidly, for a positive action particle to reverse its direction of propagation in time without reversing its energy sign (from a bidirectional viewpoint) and without immediately annihilating back to the vacuum.

It should be clear in this context that the condition of continuity of the flow of time cannot *alone* be invoked for requiring that a particle reverses its energy sign when it reverses its direction of propagation in time (so that its sign of action would remain invariant), despite the fact that such a condition does require that a particle that reverses its direction of propagation in time always retains the sign of its non-gravitational charges (so that those charges appear to be reversed from a unidirectional viewpoint), as I explained above. This is certainly acceptable given that, the postulated invariance of the sign of charge under a reversal of the direction of propagation in time is what allows the existence of the time direction degree of freedom to have physical significance (from the viewpoint of unidirectional time) when a reversal of time is combined with a reversal of energy (which leaves the sign of action invariant), while a reversal of the direction of time that leaves energy invariant is always physically significant from a gravitational viewpoint. What this means is that the creation and the annihilation of pairs of opposite action particles propagating in opposite directions of time is not independently ruled out by the condition of continuity of the flow of time.

It is now possible to understand the significance of the remarks I originally made to the effect that there could be departures from the rule enunciated in the preceding chapter that a particle cannot reverse its direction of propagation in time without also reversing its energy. Indeed, this principle was formulated under the assumption that it may no longer be valid under the very unusual conditions where the energy of matter is sufficiently large that gravitation is no longer negligible at the elementary particle level (so that the indirect gravitational interactions between opposite action particles could allow the forbidden transmutations despite the absence of contact between those particles). This exception to principle 10 would, in the context of the preceding discussion, constitute an actual requirement given that it is needed to restore the symmetry of our universe under a reversal of time. It is therefore possible to independently confirm that the whole explanation for the absence of creation of matter out of the vacuum which is embodied in this tenth principle is fully appropriate even in the context where the constraint it expresses is mostly of a practical nature and does not constitute an absolute

requirement that would be valid under absolutely all circumstances.

But what is remarkable is that despite the conditions of very short duration and high energy under which the creation of pairs of opposite action particles can be expected to have occurred, it seems appropriate to assume that the requirement of continuity of the flow of time along an elementary particle world-line would still be obeyed, given the very possibility that this assumption offers to solve the problem of the asymmetry between matter and antimatter, which occurs in the context where the distinction between those two forms of matter originates from the existence of a fundamental time direction degree of freedom whose preferred direction could otherwise have been related to the direction singled out by the thermodynamic arrow of time.

2.11   Black hole entropy

We are now entering the realm of a more uncertain domain of scientific inquiry where classical gravitation theory reaches the limits imposed by quantum indeterminacy. In order for the following discussion to be meaningful it will first be necessary to recognize that the theoretical justifications and the indirect evidence for the existence of black holes is sufficiently well established that these objects can be considered legitimate subjects of study. The objective I will try to achieve is then simply to show that it is possible to identify the degrees of freedom of matter which give rise to the exact measure of black hole entropy derived from the semi-classical theory of black hole thermodynamics. This explanation will be based on the results achieved in the previous sections while deriving an improved formulation of the discrete symmetry operations, as well as on a better understanding of the implicit assumptions entering the derivation of the semi-classical formula for black hole entropy. More specifically, I will explain that based on certain plausible hypotheses concerning the constraints that should apply on matter particles approaching a spacetime singularity, it is possible to deduce that a finite number of discrete degrees of freedom characterizes the microscopic state of the elementary particles which were captured by the gravitational field of a black hole. As a consequence, it becomes possible to actually confirm the existence of an exact relationship between those matter degrees of freedom and the binary measure of missing information or entropy which according to the semi-classical theory should be distinctive of those situations in which

event horizons are indeed present.

I will be working here under the hypothesis (now commonly recognized as appropriate) that the information concerning the matter which produced the gravitational collapse that gave rise to a black hole (or the matter which was later captured by the same object) is not lost, but is rather encoded in the detailed microscopic configuration of certain degrees of freedom associated with microscopic elements of surface on the event horizon of the object. Ignorance of this microscopic configuration when a black hole is described using the classical macroscopic physical parameters of total mass, angular momentum and charge is what gives rise to gravitational entropy. What is not fully understood presently is how we can reconcile the fact that matter appears to be characterized by physical parameters that vary in a continuous fashion, while the information contained in the microscopic degrees of freedom on the surface of a black hole must be given in binary units. What is the exact nature of the microscopic degrees of freedom of matter which would correspond with the missing information encoded in the microscopic degrees of freedom present on the event horizon of a black hole? Given the limitations imposed by the Bekenstein bound (according to which the amount of information that can be obtained concerning the microscopic state of the matter contained within any surface is also proportional to the finite number of elementary units of area on the surface) it would appear that this question actually applies to the microscopic configuration of matter under *any* condition, regardless of the strength of the gravitational field on the surface through which information about this exact state must be obtained.

It therefore seems that the problem of identifying the fundamental degrees of freedom of matter which are associated with the binary measure of entropy encoded on a two-dimensional boundary is not one that concerns only situations in which black holes are present, even though its significance is made more obvious when we are actually dealing with event horizons. I think that the fact that there is a similar measure of gravitational entropy associated with both event horizons and ordinary surfaces means that we must admit the reality of what would be occurring beyond the limits of any event horizon, despite the fact that the processes involved cannot be subject to direct observation. Thus, regardless of the practical limitations which clearly exist for actually determining the exact state of whatever microscopic degrees of freedom are to be associated with the particular measure of missing information encoded on the surface of a black hole, this problem should nevertheless be considered a tangible one, even if only because under appro-

priate conditions information about this microscopic state could be obtained. In fact, I believe that the constraints imposed by quantum theory concerning the conservation of information require that we recognize the reality of the microscopic degrees of freedom which encode all the relevant information about the matter which was captured by the gravitational field of a black hole and whose existence appears to be necessary for the consistency of the semi-classical theory.

Indeed, many recent developments confirm that contrarily to what was once argued there is no incompatibility between the general relativistic description of the event horizon of a black hole and the hypothesis that information is conserved for matter that is captured by the gravitational field of such an object. When all the dust has settled it becomes apparent that there is simply no basis to the commonly held viewpoint that the process of black hole evaporation involves fundamental, irreducible irreversibility, or that information is actually lost when a black hole decays through the emission of Hawking radiation. There is no more reason to believe that information is lost when black holes evaporate than there would be to assume that the information that appears to be lost when a drop of ink spreads into a liquid is fundamentally irretrievable. There is no rational motive for assuming that processes involving the gravitational interaction are different in this respect from any other thermodynamic process, even when we recognize that there is something more objective about the growth of entropy that is associated with the formation and the evolution of event horizons (in section 3.7 I will explain what motivates this conclusion).

What the semi-classical theory of black hole thermodynamics implies is that there does exist information about what lies behind event horizons, but that this information is missing from the description of a black hole in terms of its classical macroscopic parameters and therefore we must assume that it could only be obtained through measurements of the microscopic configuration of some physical parameters associated with the surface delimited by the event horizon of the object. The fact that a consistent theory of black hole thermodynamics actually exists means that we have no reason to expect that when such objects are involved there could be departures from the rules which govern ordinary physical systems with a large number of degrees of freedom, for which it is already recognized that any apparent information loss merely occurs as a practical limitation. In the context where it is understood that, from a physical viewpoint, information must involve a distinction, this assumption is actually supported by the existence of a relation between the

mass of a black hole and its entropy, because any distinctive features must be carried by elementary particles and when the number of particles absorbed by a black hole grows its mass necessarily becomes larger. This observation would remain significant even if it was determined that the actual microscopic degrees of freedom which are allowed to vary for matter that fell into a black hole do not consist of mere energy differences. Also, if we recognize that information, as a measure of physical distinction, can be conserved without the knowledge of some such distinction being shared by any specific observer then we are certainly allowed to assume that information persists even when black holes are involved.

Some well-known results appear to confirm that the information concerning the microscopic state of the matter which was captured by the gravitational field of a black hole may in effect be encoded in the detailed configuration of certain degrees of freedom associated with the event horizon of the object. Those conclusions are all dependent, basically, on one assumption, which is that there is a finite maximum level of accuracy applying to our description of spatial distances. This limitation would then also apply to the description of surfaces such as those which are associated with event horizons. Indeed, the still largely uncertain quantum gravitational theories which were used to achieve those results all have as a key characteristic that they involve a discrete description of physical space on the shortest scale. Based on what I have learned concerning this issue I think that I can safely argue that it is this unique particularity of current quantum gravitation theories which allows to explain that they can predict that black hole event horizons are characterized by a finite number of microscopic degrees of freedom which vary as binary parameters and which appear to encode the information about the unknown microscopic state of the matter contained within the objects. Current quantum theories of gravitation would therefore have succeeded in unveiling at least one distinctive aspect of the structure of space and its associated gravitational field in the context where quantum indefiniteness can no longer be ignored.

What was learned, more exactly, is that two events must be considered indiscernible from the viewpoint of any measurement when they would occur within intervals of space and time smaller than the natural scale of quantum gravitational phenomena. We now understand that trying to describe the state of matter and energy at a level of definition of spatial distances and during time intervals more precise than those provided by the Planck scale would constitute a superfluous characterization of physical reality. Despite

the fact that this constraint now appears clearly inescapable it is still often ignored, as when someone is talking about what may have happened at a time shorter than the Planck time after the Big Bang. Here I will assume that the limitations imposed by quantum indeterminacy, which imply the existence of a smallest meaningful spatial distance, constitute a fact which will gradually become as well established as the existence of elementary particles of matter and on which further insights can therefore be based. In such a context it would appear that if the degrees of freedom on the event horizon of a black hole are to be associated with the state of some quantum particles (perhaps gravitons) crossing this horizon then under no circumstances could two particles actually be present at the same moment in a unit of surface smaller than that which is associated with the scale of quantum gravitational phenomena. It would therefore be impossible for any physical parameter associated with such a unit of area to be attributed more than one value at any particular time (although it remains to establish what is the exact size of this fundamental unit of area and therefore it is still possible that what may now appear to be a fundamental unit of area would actually allow to encode more than one fundamental degree of freedom, as I will suggest in section 3.3).

Thus, it seems that it is from discrete elements of structure with a size of the order of the Planck interval that a proper description of the exact configuration of the microscopic degrees of freedom associated with the event horizon of a black hole can be formulated that may also be valid to some extent in the case of ordinary surfaces. What is remarkable is that it appears that the physical parameters associated with those microscopic elements of surface also vary in a discrete way, which means that they actually provide a binary measure for the entropy or missing information which characterizes those objects. Indeed, the relevant microscopic degrees of freedom on a surface can only be this or that, or yes or no, rather than assume any value from a continuous spectrum of possibilities as we go from one discrete surface element to the next. It appears that not only must we accept that space is divided in elementary units on the shortest scale, but we must also recognize that the values taken by the physical parameters associated with those discrete elements of surface can only be either one thing or another and nothing in between. Therefore, at the most fundamental level of description it would appear that the physical properties of a surface must be described using discrete elements of structure corresponding to the smallest physically meaningful measures of area to which are associated only two possible states

of some microscopic degree of freedom. In such a context the entropy of a black hole would derive merely from an absence of knowledge of the detailed configuration of this microscopic degree of freedom (characterizing elements of surface on its event horizon) which arises as a consequence of the difficulty to obtain experimental data about what is actually occurring at this level of precision of measurement.

Given the current state of knowledge concerning quantum gravity, it is not possible to determine the exact physical nature of the elementary degrees of freedom present on an event horizon, but it seems natural to assume that if a macroscopic black hole was isolated in space, then this information would have to be contained in the microscopic configuration of its surface gravitational field. In any case it is necessary to distinguish between the degrees of freedom characterizing the states of the particles which were captured by the gravitational field of a black hole and the degrees of freedom on the event horizon of the object, which merely reflect the microscopic state of the matter and which may be of a different nature from a physical viewpoint. But despite the ambiguous nature of the physical degrees of freedom which allow information to be encoded on the event horizon of a black hole it must be assumed that there exists a clear relationship between the state associated with those microscopic degrees of freedom and that of the matter from which an observer has become separated as a consequence of the presence of this theoretical boundary. What's more, given the size of the elementary units of surface on which the information concerning the microscopic state of the matter contained inside an event horizon is encoded, it appears that we would be justified to assume that the degrees of freedom of matter which we must identify are those which would apply to a description of matter at the Planck scale.

In any case I think that the existence of such a correspondence between the microscopic degrees of freedom associated with an event horizon and those of the matter it contains should be considered unavoidable even if only because we can never get more information concerning what is located beyond any surface than is obtainable by observing through this surface. But if there does in effect exist a limit to the accuracy of measurements that can be effected on a surface (due to the existence of a smallest meaningful spatial distance) then it necessarily follows that there must be a limit to the amount of information that could be obtained through a detailed probing of the processes actually occurring on that surface and this limit should naturally be expected to be proportional to the number of discrete surface elements

through which the information must flow. It should not come as a surprise, therefore, that the total area of a black hole actually provides a measure of the number of elementary units of missing information which should ultimately be related to the exact microscopic state of the matter which is located past the event horizon of the object. What's more difficult to explain is why this constraint does in effect appear to be relevant to what is actually taking place beyond event horizons rather than merely to what we can *tell* about what is going on there. Despite the enduring uncertainty associated with this question I believe that the following discussion will help clarify the nature of the relationship between the microscopic degrees of freedom on a surface and the microscopic state of the matter located within that surface.

Before I undertake the task of explaining why it is that the states of the elementary particles which have been absorbed by a black hole can become so constrained that they are allowed to match the required binary measure of missing information which is encoded on the event horizon of the object it would be appropriate to first recall what the semi-classical analysis of black hole thermodynamics has revealed. What we know in effect is that for a non-spinning black hole of mass m with an event horizon of area $A_{BH} = 4\pi R_S^2$, where $R_S = 2mG/c^2$ is the Schwarzschild radius of the black hole, the entropy is given by $S_{BH} = \frac{1}{4} A_{BH}/A_P$, where $A_P = l_P^2$ is the Planck area given in terms of the Planck length which is defined as $l_P = (\hbar G/c^3)^{1/2}$ and the units are chosen so that Boltzmann's constant k is equal to unity. In general, a black hole would therefore have an entropy that is determined by the value of the area of its event horizon in Planck units of surface divided by a factor of four. Given that entropy is simply a measure of the information that is missing from the description of a black hole in terms of its macroscopic parameters of mass, radius, or area it seems that the amount of information encoded in the unobserved microscopic degrees of freedom characterizing the surface of the object is equal to one fourth its area in natural units. It was pointed out by Gerard 't Hooft, before the previously mentioned results obtained from quantum gravity were derived, that this actually means that information appears to be encoded on the surface of the black hole in binary units corresponding to an area equal to four Planck areas.

Now, if we are willing to accept that the Planck unit of area may actually be given as equal to $A_P = 4\pi l_P^2$ (following the traditional formula for the area of a sphere in terms of its radius) when the mass of a black hole approaches the Planck mass (from higher values associated with macroscopic

event horizons) then an interesting result can be shown to follow. Indeed, using the above equation for A_{BH} we can deduce that the event horizon of what I would call an elementary black hole, with a mass equal to exactly one Planck mass $m_P = (\hbar c/G)^{1/2}$, should have an area that is actually equal to four such Planck areas. Using the formula for the entropy of a conventional black hole I would thus be allowed to conclude that the detailed configuration of the microscopic degrees of freedom on the surface of a Planck mass black hole must carry one single binary unit of information. I think that the outcome of this simple derivation is extremely significant, because on the basis of the hypothesis that there can be no significance in attributing existence to a particle which would occupy a volume smaller than that which is associated with the most elementary unit of area (as current quantum gravitational theories appear to require) it seems necessary to assume that such an elementary black hole, would be formed of at most one single elementary particle and in such a case we have no choice but to attribute the information encoded in the microscopic degrees of freedom on the surface of the black hole to its matter content.

But if, in the case of an elementary black hole at least, the missing information encoded on the event horizon of the object must definitely be associated with the single Planck energy particle it contains (which need not necessarily have a large rest mass) then even for a black hole of larger mass it should be possible to associate this binary information with the states of matter particles contained within the surface, despite the fact that according to the above equations the entropy of a black hole S_{BH} is not in general proportional to its mass m, but rather to its mass squared (so that entropy rises faster than the matter content). The fact that no simple relationship between entropy and matter content appears to exist in the general case of a macroscopic black hole is simply due to the fact that the gravitational field must itself carry a portion of the entropy when large accumulations of matter are involved. However, in the context where the particles mediating the gravitational field are to ultimately also be understood as being a form of matter we would have no choice but to associate the entire amount of missing information associated with a black hole's event horizon with the 'matter' content of the object, which would then include gravitons. In any case, if an elementary Planck mass black hole, containing a most elementary particle with an energy of the order of the Planck energy, can be associated with the smallest unit of information then it requires that we recognize that the binary nature of the microscopic degrees of freedom on the event horizon of

any black hole is a reflection of the existence of states of matter which can only vary in a discrete way.

So, what are exactly those degrees of freedom prevailing for matter trapped by the gravitational field of a black hole? When we ask this question in the context where the information associated with an *elementary* black hole is understood to provide a complete description of the state of a most elementary particle in the conditions where an event horizon is constraining the motion of this particle it appears necessary to assume that its state must be completely definable by one single binary unit of information. It may therefore appear that we should be seeking to identify a unique physical parameter that reverses under a given discrete symmetry operation as being the binary degree of freedom related to the information encoded on the event horizon of our elementary black hole. But if we are to assume that the same fundamental parameters characterize the spacetime-related properties of matter under all conditions then it rather seems that all the truly independent discrete symmetry operations, like the previously defined T, P, and M operations should have their counterpart in the information associated with the state of the particle forming an elementary black hole. Indeed, all of those reversal operations allow to distinguish the sign or the direction of some physically significant property of elementary particles and there is no a priori reason why only a subset of those variable properties should need to be taken into account in the characterization of the discrete degrees of freedom applying at a fundamental level in the presence of an event horizon.

It must be clear that if all of the independent discrete symmetry operations were considered to determine one distinct degree of freedom of a particle confined by the event horizon of an elementary black hole then we would need not one binary unit of information or one bit to be encoded in the microscopic configuration of the gravitational field on the surface of the object, but rather three bits. Indeed, with two yes or no questions we can determine the action sign preserving direction of time intervals (reversed by T or not reversed) and the action sign preserving direction of space intervals (reversed by P or not reversed), which already allows to distinguish four states of matter (identity being the state where neither space nor time is reversed). The distinctions which exist between each of those four states as they appear from the bidirectional and the unidirectional time viewpoints are illustrated in figures 2.1 and 2.2. With an additional yes or no question we can then determine the sign of energy or action (reversed by M or not reversed), which doubles the number of states of matter that can be dis-

tinguished, so that we can differentiate between the eight possible states of matter related by the discrete symmetry operations defined in the preceding sections. The C symmetry operation being a combination of T and P does not provide an additional distinct degree of freedom and therefore need not be considered here (even though we may as well consider only T and C or only P and C to provide the relevant discrete degrees of freedom and then it would be P or T which could be ignored). But three bits is not equal to one bit and so it may seem that there is a problem with associating the missing information encoded on the surface of a black hole with the degrees of freedom transformed by the discrete symmetry operations, despite the fact that those parameters should in effect characterize the states of elementary particles under all circumstances.

However, I believe that this discrepancy cannot be assumed to rule out the validity of the theoretically unavoidable conclusion that any binary distinction between the states of the matter particles that crossed the event horizon of a black hole must be a reflection of the structure underlying the previously defined discrete symmetry operations which together allow to transform all physically meaningful states of matter that vary in a binary way. I will show that very restrictive constraints actually limit the variability of certain microscopic physical parameters whenever black holes are involved. Those limitations imply that some parameters which may otherwise appear to be independent from one another actually vary together when subjected to various reversal operations. Some microscopic physical parameters are also restricted to a subset of the values they would otherwise be allowed to take. This actually contributes to reduce the number of binary units of information needed to specify the microscopic states of particles trapped by the gravitational field of a black hole. A further insight will be needed, however, to allow the number of binary units of information required for achieving this complete description of the state of gravitationally collapsed matter to be made entirely compatible with the measure of black hole entropy derived from the semi-classical theory.

In order to clarify the situation regarding what variations are allowed for the various microscopic properties of elementary particles when matter has become confined by the gravitational field of a black hole we may first recall that the three macroscopic physical parameters characterizing a black hole are its total mass m, its total charge q, and its overall angular momentum j. To those three parameters I would add the momentum p, which is not

usually considered to define the macroscopic state of a black hole, but which I believe provides essential information required to identify the parameters which must be taken into account in defining the microscopic state of the particles that form such an object. It must be clear that each of those macroscopic parameters must be allowed to vary not just in magnitude, but also in sign or in direction. The total mass m in particular must be conceived as being either positive or negative depending on the overall sign of energy of the black hole. This is also an aspect that is usually not taken into consideration in the conventional treatment of black hole thermodynamics, but which must be recognized as a necessary assumption in the context where the existence of negative energy matter is theoretically unavoidable.

A different question would be to ask whether the sign of energy or action is a variable parameter for the particles forming a black hole. Given that I have already argued that negative energy matter cannot be absorbed by a positive energy black hole, it may seem that only positive energy states need to be taken into account in describing the microscopic configuration of the matter that was captured by the gravitational field of a positive mass black hole. It is important to understand, however, that one cannot assume that all black holes with a given mass sign must at all times be formed only from particles with the same mass sign as that of the object itself, because even if no particle of energy sign opposite that of a given black hole can cross its event horizon from the outside, the possibility that a positive energy black hole may already contain negative energy matter (or vice versa) is very real and must be taken into consideration. Indeed, it is indisputable that a positive mass black hole with a very large radius and a rather low density could potentially form despite the initial presence of some comparatively small amount of negative energy matter inside the surface that is to become its event horizon. Thus, it is not strictly forbidden for a positive energy black hole to contain negative energy matter even though this matter would only be allowed to be present inside the event horizon associated with such a black hole if it was already contained inside the surface that became this event horizon before the gravitational collapse occurred.

But, even if a positive energy black hole was to contain negative energy matter, this matter would not remain in this situation for very long, because it would rapidly be expelled by a gravitationally repulsive force equivalent in strength to that which is attracting the rest of the matter toward the central singularity, so that the black hole would actually end up containing exclusively particles having the same sign of energy as its own. Furthermore,

even if the sign of energy of the particles contained within any surface was to constitute a relevant microscopic degree of freedom (transformed by the M symmetry operation) that could contribute to the measure of information encoded on this surface, there is an independent motive for assuming that black holes are composed of matter with a sign of energy that is necessarily the same as that of their own total mass. Indeed, if I want to explain the results of the semi-classical theory of black hole thermodynamics I have no choice but to first assume that the energy sign of every matter particle forming a black hole corresponds to the energy sign of the object itself, because the conventional theory is based on the implicit hypothesis that positive energy black holes exist in a stable state and are not in the process of releasing negative energy matter, which means that they must be formed exclusively of positive energy matter. I will therefore assume as a first approximation that a knowledge of the sign of mass of a macroscopic black hole allows to determine the energy signs of *all* the matter particles whose states are reflected in the detailed configuration of the microscopic degrees of freedom on the event horizon of the object.

It should be clear that under such conditions it cannot be assumed that the energy sign of particles, which is transformed by the action sign reversal symmetry operation M, constitutes the one binary degree of freedom per elementary unit of area which is associated with the measure of black hole entropy provided by the semi-classical theory, because if that was the case then in the most common of situations nearly all the microscopic physical parameters of a black hole would be fixed by a knowledge of the sign of mass of the object and no information would be missing from the macroscopic description. It would thus follow that entropy would always be minimum, which is certainly not desirable given that the semi-classical theory rather requires entropy to be maximum when matter collapses into a black hole. The constraint imposed by the sign of mass of a black hole on the energy sign of its constituent particles may not be so significant, however, given that even if we ignore any additional degree of freedom which could be associated with the other discrete symmetry operations, a determination of the sign of energy cannot alone be considered to exhaust the requirements of a complete description of the state of the matter particles forming a black hole, because in principle energy must also be allowed to vary in magnitude.

For now, we may choose to leave aside that difficulty, but then we are still left with having to explain how it can be that the other two independent discrete symmetry operations which should also characterize the states of

matter under all conditions provide at most only one single binary unit of information, even though they together transform two degrees of freedom. As I suggested above, those two symmetry operations may be chosen to be the action sign preserving time reversal operation T and the action sign preserving space reversal operation P. You may recall that in the context of the redefinition of the discrete symmetry operations which I proposed in a previous section the T symmetry operation must be assumed to reverse all momenta, as well as all angular momenta and all non-gravitational charges, even if merely from the unidirectional time viewpoint. The P operation on the other hand has absolutely no effect on the direction of angular momentum or the sign of charge, from any viewpoint, but must be considered to reverse the direction of momentum and the handedness of particles (as indicated in table 2.4). Thus, taken together the T and P symmetry operations would affect all the physical parameters defining the microscopic states of particles which add up to produce the total momentum p, angular momentum j, and charge q parameters that characterize the macroscopic properties of a black hole with a given sign of mass.

Yet this does not necessarily mean that all that must be specified to determine all of those macroscopic physical parameters are the signs of the microscopic parameters transformed by the T and P operations (reversed or not reversed for each of the two symmetry operations) for every elementary particle that forms a given black hole. There is, in effect, no a priori reason to assume that the momentum of elementary particles (like their energy) can vary only in sign and it would rather seem that not only must this parameter be allowed to vary in magnitude like energy, but it must in addition be allowed to vary in direction, not in a binary way like the sign of energy, but as a continuous two-dimensional angular variable, which would forbid its complete determination through a knowledge of the value that would be taken by one single binary degree of freedom. What's more, even if under ordinary circumstances an action sign preserving reversal of time intervals generated by T would affect the direction of angular momentum (because it would reverse momentum independently from position) it would not affect the handedness of particles and in the context where we are trying to identify the microscopic configuration associated with the states of elementary particles present at the Planck scale it appears necessary to restrict our account of the spin state of matter particles to handedness. But while an action sign preserving reversal of space intervals obtained by applying P would actually reverse the handedness (because it would reverse the momentum of particles

without also reversing their spin), we have no reason to assume that the spin could not itself reverse independently from momentum, thereby also reversing the handedness. It would then appear necessary to specify the handedness of particles independently from the other degrees of freedom which are reversed by those two symmetry operations. As a consequence, only the sign of charge of a given particle can be assumed to be entirely determined by its dependence on the redefined time reversal symmetry operation T when the effects of such a transformation are considered from the unidirectional time viewpoint, which usually applies in a classical context.

Now, despite the fact that the T operation reverses both momentum and charge, it certainly seems appropriate to assume that as far as those microscopic physical parameters are concerned we are actually dealing with two distinct degrees of freedom, because momentum can also be independently reversed by the P operation. But even though it may appear obvious that the sign of charge should be independent from the direction of momentum it is reassuring to observe that from a bidirectional time viewpoint this hypothesis is unavoidable given that the variation of the sign of charge only occurs from a unidirectional viewpoint and is actually the consequence of a reversal of time intervals obtained while leaving the sign of action invariant, which would in effect reverse the sign of energy, but leave invariant the direction of momentum. In any case the outcome of this reflection is that we have to accommodate three independent microscopic degrees of freedom which are the sign of charge, which is reversed by T, the direction of momentum, which is reversed by P, and the handedness of elementary particles, which can be reversed independently from charges and momenta. The variation of any other physical parameter (except for the sign of action) can then be derived from a knowledge of the variation of those three independent parameters. It is also important to mention that despite the fact that what I'm seeking to determine are the degrees of freedom which would apply on a very small scale at which the fundamental interactions would presumably be unified, I'm nevertheless assuming that the sign of any non-gravitational charge would remain a parameter distinct from the sign of action (the gravitational charge), because the variation of the sign of charge would here occur merely as a secondary consequence of a reversal of the direction of propagation in time, which must still be considered a significant change clearly distinct from a reversal of action (which also involves a reversal of time, but which leaves the sign of energy unchanged), even under such conditions.

If we recognize the appropriateness of those remarks it would then seem

that the situation may be even more problematic than I indicated above, because despite the fact that we are considering as relevant only those parameters which are affected by the redefined discrete symmetry operations, the degree of freedom associated with handedness would provide an independent contribution (dependent on the direction of spin) to the measure of missing information concerning the microscopic state of the matter that crossed the event horizon of a black hole. This contribution would add to those provided by the degrees of freedom associated with the sign of charge and the momentum of a particle (which from a bidirectional time viewpoint are dependent on the sign of time intervals and the sign of space intervals, respectively). It would then seem that we still need at least three binary units of information to completely describe even just the *signs* of all the relevant physical parameters characterizing the microscopic state of matter under such conditions. But, as I will explain, the existence of an independent degree of freedom related to handedness, far from creating a problem is in fact precisely what allows the appropriate measure of entropy to be derived in the presence of an event horizon.

It is while I was trying to visualize what would happen to a negative mass body which would find itself inside some surface that was about to become the event horizon of a positive mass black hole that I realized that both positive and negative mass particles would actually be submitted to very restrictive constraints when experiencing the effects of the gravitational field which exists inside the region delimited by the event horizon of a black hole. Indeed, a negative energy particle which would happen to be located near the center of a positive mass black hole at the time of its formation would immediately be repelled outward by a force as large as that it would experience inside the most powerful of particle accelerators. While it is being ejected outside the event horizon, the negative energy particle would reach an arbitrarily high (negative) energy and its momentum would also become arbitrarily large in the direction opposite the forming central singularity (considered as the point where the density of the dominant form of energy reaches its theoretical limit), regardless of what its initial state of motion was. The nearer to the center of the object the particle would initially be, the larger its final negative energy would be when it would emerge from the event horizon of the positive mass black hole. But given that the force which accelerates the particle is always directed away from the forming singularity, it follows that the components of the momentum in any other direction would become com-

pletely negligible in comparison with the component directed away from the singularity. In fact, if we were to consider only particles literally emerging from a positive mass singularity, we would end up (in the absence of collisions with infalling positive energy matter) obtaining particles reaching the event horizon with a maximum (negative) energy and a momentum invariably directed along the positive normal to the surface of the black hole. In other words, we would always obtain (in the absence of interference) particles in a very specific state of motion.

The process would be even more constraining for a positive mass body in the gravitational field of a positive mass black hole given the rising tidal effect which in this case compresses the object laterally and stretches it vertically as it is accelerated in the direction of the singularity. In such a case we would necessarily end up with a very focused beam of particles whose lateral motions would again be completely negligible. Indeed, the force attracting the particles toward the singularity of the black hole would grow with time from the moment they cross the event horizon, eventually becoming so large that the energy of the particles would become as high as it can be, while the horizontal components of their momenta would become completely negligible in comparison with the vertical component of their momenta oriented toward the central singularity. Any residual lateral motion would simply contribute to increase or decrease the total angular momentum of the black hole whose rotation is shared by all the particles that fell into the singularity (as a consequence of collisions and relativistic frame dragging) and should not therefore contribute to entropy (as a measure of missing information concerning microscopic degrees of freedom). Thus, when a positive energy particle reaches the singularity of a positive mass black hole its momentum (in the frame of reference relative to which the object is not rotating) has basically become a unidirectional variable. In fact, space itself must be considered to become analogous to unidirectional time for a positive energy particle that crosses the event horizon of a positive mass black hole, but what I came to understand is that this actually means that momentum would then become a fixed parameter with a unique direction and a maximum magnitude. As a result, we once again obtain a unique final state of maximum energy and invariant momentum.

The crucial assumption in the present context is that a maximum energy must actually exist. I believe that this conjecture is appropriate given that in a quantum gravitational context the existence of a minimum meaningful time interval or spatial distance implies the existence of a similar limit for

the magnitude of energy. Indeed, if we are considering the state of a particle that occupies a region of space of the order of that which is associated with the smallest physically meaningful spatial distance (the Planck length), then upon reaching an energy of the order of the Planck energy the particle would itself become a black hole. Thus, if the particle was to gain even more energy the area of the event horizon enclosing it would simply grow to encompass a larger region, which we could only associate with the presence of a larger number of elementary particles instead of assigning the original single particle with a larger energy. Hence, there would be no sense in attributing one single elementary particle at the Planck scale with an energy larger than the Planck energy. The situation we encounter here is somewhat similar to that which we have in quantum chromodynamics, where beyond a certain threshold the energy spent at trying to separate two oppositely charged quarks in a meson no longer contributes to increase the distance-dependent attractive force between the two quarks, but merely end up splitting the original particle into two new mesons thereby neutralizing the force that existed between the original two quarks.

I would therefore suggest that we assume that the particles that reach a singularity after having been accelerated by its gravitational field must be in a state of maximum energy which we must recognize as the Planck energy. Given that it is not that difficult to visualize what would happen to a positive energy particle which would cross the event horizon of a positive mass black hole it is surprising that it had never been fully realized what the outcome of such a process would mean for any description of the final state of a gravitational collapse. But I believe that it is crucial to recognize, in order to clarify the whole question of black hole entropy, that what happens when matter collapses to form a black hole is that it invariably reaches a state in which every particle has a Planck energy and a correspondingly large momentum characterized by a unique invariant direction which is straight toward the singularity, regardless of the initial motion of the particles at the time when they crossed the event horizon of the black hole.

Here it must be understood that despite the fact that the wavelength of the light emitted by a positive energy particle which is about to be absorbed by a positive mass black hole would be infinitely redshifted (from the viewpoint of a remote observer not moving with respect to the event horizon of the object) and would show time as standing still, we are nevertheless allowed to assume that the events occurring after the particle crosses the event horizon of the black hole can be characterized in a physically meaningful way. It is

certainly appropriate to consider, in particular, that a particle's momentum will indeed keep increasing in the direction toward the singularity, as I'm suggesting would be the case, because time dilation does not mean that the particle itself would become motionless, but merely that the signals it emits are infinitely redshifted by the gravitational field of the black hole (still from the viewpoint of a remotely located, motionless observer). Thus, despite the fact that signals would show the particle as apparently immobilized on the event horizon we must still assume that this particle actually crosses the event horizon and in a finite time acquires an energy which relative to a motionless outside observer would be arbitrarily high.

Also, the notion that, from the viewpoint of an external observer, a positive energy particle could in fact acquire a negative energy after it crosses the event horizon of a black hole would only be appropriate if we were to consider that the negative gravitational potential energy reduces the energy of the particle itself into negative territory. But in fact this is not an appropriate approach to defining the energy of matter (especially in the context where the true properties of negative energy matter are understood to make such a notion implausible) given that as far as this potential energy is concerned we are actually dealing with a distinct contribution to energy which is that of the gravitational field. The truth is that the kinetic energy of the particle itself would keep increasing to arbitrarily large values even if this energy is compensated by a growing negative contribution to the energy of the gravitational field associated with the interaction of this particle with the rest of the matter in the black hole.

However, one may perhaps question the conclusion that momentum would have a fixed magnitude for any positive energy particle that reaches the singularity of a positive mass black hole in the context where the rest mass itself may be a variable parameter. It is true in effect that the magnitude of this momentum would depend on the rest mass of the particle which is accelerated in the gravitational field of the black hole given that all masses have the same acceleration and are therefore subjected to the same velocity increase. But in the context where we are dealing with final kinetic energies which are so large it appears appropriate to assume that the energy associated with the rest mass of the particles which are reaching a singularity after having been absorbed by a black hole would be negligible or null, so that if the total energy of those particles is the Planck energy (the maximum physically meaningful measure of energy) we are allowed to conclude that the final magnitude of their momentum would always be what we may call

a Planck momentum, understood as the maximum theoretically meaningful value of momentum which can be carried by a massless particle (associated quantum mechanically with the smallest meaningful measure of spatial distance). Under such conditions we would have no choice but to recognize that the final magnitude of the momentum of all particles reaching a black hole singularity should actually be an invariant property, just like the direction of this momentum.

Now, I initially thought that it would be appropriate to assume that if space actually comes to an end for matter that reaches a singularity, then momentum, as the conjugate attribute to space, simply cannot continue to evolve after the final stages of a gravitational collapse. But some relatively recent results from loop quantum gravity appear to show that the final state of a gravitational collapse is not a singularity, but merely a state of maximum matter density which would immediately be submitted to a 'quantum bounce' that would turn the collapse into a process of outward expansion. It is sometimes argued that this might be problematic given that if a black hole was to expel matter it seems that entropy could decrease in the process. However, given that black hole evaporation does involve a *local* decrease of entropy for the black hole itself (independently from its environment) over its entire lifetime, then the prediction that the singularity would decay may not be as paradoxical as one might assume. In fact I think that if black holes do evaporate, then something like the quantum bounce must occur, so that there remains no singularity in the final state, when the mass of the black hole itself has become minimal. The perceived problem merely arises when we fail to recognize that the near infinite time dilation that is attributable to the enormous gravitational field of a black hole implies that the process of gravitational collapse and the following quantum bounce that would take place in a finite and relatively short time from the viewpoint of the matter that falls toward the singularity, actually appear to occur over the entire lifetime of the object from the viewpoint of an external observer.

In fact, it seems that all the matter that ever crosses the event horizon of a black hole actually reaches the maximum density state at nearly the same time (at which point all matter with an energy sign opposite that of the black hole has already been expelled outside the event horizon), but from the viewpoint of an external observer this whole process, as well as the quantum bounce that follows it, would take place over the arbitrarily long period of time during which the black hole would exist. The quantum bounce, if it could be observed from outside the event horizon, would therefore appear as

a slow process whereby all the energy of the matter that fell toward the singularity would be released in the form of thermal radiation produced by the black hole while it evaporates. Indeed, given that the gravitational entropy is maximum for matter that collapses to a generic future singularity (as will be emphasized in sections 3.7 and 3.8) it follows that any information about the structure of the objects which were absorbed by the black hole would be lost from a *practical* viewpoint, even though it would later be released in high entropy form as the black hole decays through the emission of thermal radiation, and therefore no violation of the second law of thermodynamics would be observed.

Despite what is sometime suggested, therefore, the process that takes place following a generic quantum bounce is different from a white hole (conceived as the time-reverse of a black hole gravitational collapse), because the matter which is released following the bounce has high entropy and does not consist in the same macroscopic objects that originally fell through the event horizon. Yet it must also be the case that the particles which are released by a black hole as Hawking radiation actually consist in the same matter that initially fell through its event horizon. It may well only be the widespread ignorance of the unavoidable character of this interpretation that prevents us from acknowledging the fact that the information about the matter that was absorbed by a black hole is not really lost as a result of the evaporation process, but is actually contained in the detailed microscopic state of the emitted radiation.

In any case, if we are willing to recognize that the description provided by current quantum theories of gravitation constitute the most accurate account of the process of black hole gravitational collapse that one can derive, it would follow from the preceding analysis that over the entire lifetime of a black hole the particles with the same energy sign as that of the object would actually be either collapsing, with maximum momenta directed toward the singularity, or expanding, with maximum momenta directed in the exact opposite direction (as would occur after the quantum bounce takes place). This is because the time dilation effect is indeed maximum when the particles are near the singularity and are either all collapsing with maximum energy or all expanding with maximum energy, so that from an external viewpoint they would appear to spend most of their time in either one of those two states. Those discrete states would therefore be the ones that need to be reflected in the configuration of the microscopic degrees of freedom on the event horizon of the black hole before the object actually evaporates to nothing. More

specifically, the detailed configuration of the microscopic degrees of freedom on the event horizon of a black hole must be considered to reflect the state of the matter it contains at the time immediately before or immediately after it reaches the singularity of the object.

If we agree on the plausibility of the above conclusions concerning the final state of any particle involved in a gravitational collapse, then we need to recognize that the consequences of the assumption that the sign of mass of a stable-state black hole would determine the sign of energy of each of its constituent particles are much more dramatic than one may have expected. Indeed, it now appears that not only must the signs of energy of the particles in the final state of a gravitational collapse be considered to be completely determined by a knowledge of the sign of mass of the object, as I already suggested, but the magnitude of those energies is also to be considered an invariant parameter, which therefore cannot contribute to the entropy of the black hole. What's more, a similar conclusion applies for the momenta of the particles present in the final stages of a gravitational collapse which must be considered to be completely determined not just in magnitude, but also in direction, once the sign of mass of the black hole is known. Therefore, the momenta of the particles which crossed the event horizon of a black hole, like their energies, cannot contribute to the measure of entropy or missing information which determines the temperature of the object. Only in the presence of negative energy matter would the direction of the momentum of particles be allowed to vary (in sign) at any specific time inside the region delimited by the event horizon of a positive mass black hole. Indeed, even if only one magnitude of energy can be considered significant for matter located inside the event horizon of a black hole this would not exhaust the number of possibilities regarding the momentum states of particles in the context where the energy could actually be either positive or negative, because the direction of momentum would then depend on the sign of energy of the particles. But if we are to concentrate on accounting for the microscopic configuration of black holes which have reached a stable state (from an external viewpoint) then this possibility can indeed be ignored.

As a matter of fact, if the only parameters relevant to a description of the microscopic state of the matter encoded on the event horizon of a black hole were the signs of energy (actually the signs of action) and the directions of momentum of elementary particles, then given that only one magnitude of energy is allowed for the matter particles that reaches a singularity, we would have to conclude that stable-state black holes have minimum entropy,

because they only contain matter of one particular energy sign which also happens to determine the direction of all the momenta. Indeed, for a stable-state black hole, the microscopic state of energy and momentum of all the matter particles can be fully determined merely by providing the *sign* of energy of the black hole. Under such conditions there would be no meaning in trying to associate the energy sign or momentum direction degrees of freedom with some measure of entropy that would be significant from a thermodynamic viewpoint. Yet given that the case of stable-state black holes is more representative of the situation we have in practice when event horizons are actually present, I would argue that this difficulty does not mean that such black holes cannot be used to derive very general results, but rather that it is necessary to recognize the relevance of additional degrees of freedom also transformed by the discrete symmetry operations. In other words, the microscopic degrees of freedom of matter which are reflected in the detailed configuration of the gravitational field on the event horizon of a stable-state black hole simply cannot be those which are associated with the energies and the momenta of the particles that collapsed to form the object.

I'm therefore allowed to conclude that it must be the remaining degrees of freedom, which I previously identified as being the sign of charge and the handedness of matter particles, that would freely vary for particles in the final stages of a gravitational collapse and in such a way potentially contribute to the measure of missing information concerning the microscopic state of a black hole. Given that momentum direction itself is fixed, it seems that the handedness of particles could in effect allow to determine one binary degree of freedom which would vary upon a reversal of the direction of spin. Indeed, when the direction of momentum is fixed, the variable direction of spin relative to this momentum direction is the only parameter that can still vary. But then, what about the contribution by the sign of charge of those particles that crossed the event horizon of a black hole? Shouldn't this free parameter also contribute to the measure of entropy derived from the semi-classical theory of black hole thermodynamics? Even if we assume that there is only one type of charge for the elementary particles present at the unification scale, certainly information should be needed to specify whether this charge is positive or negative, given that charge appears to reverse when the direction of time intervals is itself reversed.

I have explained why one binary unit of information would be enough to account for all but one of the fundamental degrees of freedom of any positive energy particle present in the final stages of a gravitational collapse, but it

would seem that another bit is required in each corresponding elementary unit of area on the event horizon of the associated black hole to determine either the handedness of particles or the sign of charge associated with the direction of propagation in time of those particles, which clearly varies independently from that of momentum (which is associated with the direction of propagation in space). We have gone from four bits to two bits per unit of area, but that is still one bit away from the single bit that the semi-classical theory of black hole thermodynamics indicates must be encoded in the detailed configuration of the gravitational field on the surface of an elementary black hole. Given that what I'm seeking to allow is a complete determination of all the physical properties of the particles present inside the event horizon of a black hole from a knowledge of the value of all the relevant discrete degrees of freedom it would seem that I have fell short of this objective. I would like to suggest, however, that in fact the problem we seem to have encountered is not real.

The truth is that there is no contradiction between my account of the quantity of information required to completely describe the state of an elementary particle which was captured by the gravitational field of a black hole and the measure of missing information encoded in the microscopic state of the gravitational field on the event horizon of such an object at any specific time. To understand what motivates this conclusion we must first acknowledge that the formula for black hole entropy was derived from arguments related merely to the thermodynamic properties of the gravitational field itself and only in the context where the measure of missing information involved must be used in setting the strictly thermodynamic relationships between quantities like the surface gravitational field and the temperature of the thermal radiation emitted by a black hole. But, if it is indeed the case that only one out of two bits concerning the state of matter contained within an elementary black hole is encoded in the detailed configuration of the gravitational field on its surface, I think that this is because there is more information encoded in some other physical properties of black holes that do not contribute to the measure of entropy provided by the semi-classical theory of black hole thermodynamics and associated merely with their surface gravitational fields.

Once we have recognized that there must be more information concerning the microscopic state of matter contained within a surface than is provided by the detailed configuration of the gravitational field on that surface, what becomes crucial to understand is that there is no reason to assume that

the gravitational field should provide information about the microscopic distribution of some non-gravitational charge, because that information must actually be contained in the detailed microscopic configuration of the field of interaction associated with this particular charge. It is surprising in fact that this requirement was never considered before, because when one carefully thinks about this question, it is hard to arrive at a different conclusion. If the missing information about energy, momentum and angular momentum (as the physical properties of particles which constitute the source of gravitational fields) is to be associated with the microscopic state of the gravitational field then it is also quite unavoidable that missing information about, say, the electric charge is to be associated with similar microscopic aspects of the electromagnetic field. There is in fact absolutely no reason to assume that the detailed configuration of the electric charges which are the source of the electromagnetic field should be determined from information contained in a different force field, which would here be the gravitational field. It must be clear, however, that any information associated with the electromagnetic field on the surface of a black hole that would encode the details of the configuration of electric charges inside the object would have to be contained in the *microscopic* degrees of freedom of the electromagnetic field and would not be reflected in the classical macroscopic parameters of this field, which means that the hypothesis that black holes have no 'hair' would still be valid.

Now, if the missing information concerning the microscopic distribution of electric charges or electric charge signs inside a given surface (whether or not this surface is that of a black hole) can only be encoded in the detailed configuration of the electric field on the boundary delimited by that surface (even when the total charge inside the surface would be null) rather than in the configuration of the gravitational field on the same boundary, then it means that a theory that would seek to derive a measure of the amount of information necessary to determine the state of the matter contained inside this surface based only on features of the gravitational field present on the surface (which in the case of black holes would be the event horizon) would necessarily fall short of providing the accurate value. Therefore, the results derived from the semi-classical theory of black hole thermodynamics concerning the relationship between the entropy of a black hole and the area of its event horizon (considered as a property of the gravitational field) would not rule out the existence of an additional amount of missing information associated with the exact microscopic state of the matter trapped within such

a surface.

I believe in effect that the missing information concerning the sign of charge of every particle forming a black hole, which is transformed by the T symmetry operation (from a unidirectional time viewpoint), cannot contribute to the measure of disorder or entropy associated with the gravitational field of the object and this explains that it need not be taken into consideration when deriving the statistical mechanical properties of black holes associated with the various properties of their event horizons[4]. This is why we were allowed to ignore the existence of this information when elaborating the semi-classical theory of black hole thermodynamics from which the conventional measure of black hole entropy was derived. It thus appears that one additional binary unit of information (distinct from that which must be associated with handedness) is indeed missing concerning the state of every elementary particle in the final stages of a gravitational collapse. This information would allow to determine the sign of charge of each and every particle which contributes to fix the total charge q of a black hole, or more specifically the direction of time intervals along which those particles are propagating and from which depend the sign of their charges from a unidirectional time viewpoint. We are then allowed to assume that this is the binary unit of information which is actually associated with the T symmetry operation (or alternatively the C symmetry operation) defined in a previous section.

It would therefore seem that there is in effect more information associated with the microscopic state of the matter contained in a black hole than is encoded in the detailed configuration of the discrete gravitational field degrees of freedom present on the event horizon of the object. But I have explained why we should not expect this missing information to contribute to the conventionally derived measure of black hole entropy. Instead, the additional information should be associated with the entropy contained in the interaction fields associated with the distribution of non-gravitational charges, which would give rise to its own independent contribution to the temperature of a black hole. In this context it is important to note that there actually exists an analogue to the Hawking radiation process associated

---

[4]This conclusion is especially appropriate given that from the bidirectional viewpoint of time it is the sign of energy that reverses under application of a T symmetry operation and given that this reversal is combined with a reversal of physical time intervals, then it is not significant from a gravitational viewpoint and therefore it should not be reflected in the microscopic properties of the gravitational field on the event horizon of a black hole.

with the gravitational field of black holes and which involves the electromagnetic field. It is a known fact indeed that, past a certain magnitude, the electrostatic field surrounding a charged nucleus would induce pair creation processes similar to those associated with the radiation emitted by a black hole and I believe that this phenomenon would allow a similar treatment of the thermodynamic properties which according to the above proposal should be associated with any distribution of non-gravitational charge. Only, in the case of non-gravitational charge we are usually dealing with situations where the total charge is indeed null even when large amounts of positive and negative charges are present inside a surface. Such situations are therefore more analogous to that which is occurring when the measure of gravitational entropy is constrained merely by the Bekenstein bound and both positive and negative energy matter are present together inside a surface.

If you have understood the essence of my argument, then there should be no doubt that the only missing information which is actually encoded in the microscopic configuration of the degrees of freedom associated with the surface gravitational field on the event horizon of a black hole is that which allows to determine the handedness of every particle it contains using merely one single bit of information for every elementary particle. This conclusion should perhaps have been expected given that the direction of spin is the only physical parameter that reverses only when a reversal of the sign of action M is applied, but which is nevertheless allowed to vary for particles with a specific sign of energy submitted to maximally strong gravitational fields, such as those present in the vicinity of a black hole singularity. In any case, if we are willing to accept the validity of the arguments on which this deduction is based it would then follow that we now have an explanation not only for the fact that the states of the matter particles trapped by the gravitational field of a black hole vary as discrete variables, but also for why only one such variable (instead of three of four) actually contributes to the measure of missing information which must be taken into account in determining the thermodynamic properties of such an object associated with its surface gravitational field. As a result, the measure of information associated with the matter content of an elementary black hole is allowed to match the value of entropy derived from the semi-classical theory of black hole thermodynamics, which requires each elementary unit of surface (equal to four Planck areas) to encode one binary unit of information.

Therefore, it is now actually possible to at least confirm the existence of a definite relationship between the microscopic state of the quantized gravi-

tational field on the surface of a black hole and actual states of the matter it contains. What held the key to a better understanding of the exact nature of the degrees of freedom characteristic of the states of matter submitted to a gravitational collapse was the recognition that for matter particles reaching a black hole singularity the only relevant variables are the signs of all those physical parameters which are transformed by the previously discussed discrete symmetry operations. It is remarkable that the sign of handedness in particular should be one of the only fundamental parameters of elementary particles (along with the sign of charge) that is not constrained to any specific value by the conditions prevailing in the final stages of collapse into a spacetime singularity and that it must therefore alone contribute to the measure of entropy associated with the gravitational field of a black hole. This is certainly the most significant outcome which has emerged from my re-examination of the question of discrete symmetries as it arises in a semi-classical context.

If we now return to the more general case for which the density of matter is not large enough to produce an event horizon and the possibility for positive and negative action matter to be present together inside a surface cannot be ignored, it transpires that this is a situation in which more information would be required to describe the microscopic configuration of matter, because more states of motion are allowed for the particles in the period before such a configuration reaches a stable state. Indeed, even when an event horizon associated with a positive mass black hole is present it is clear that while a positive energy particle would be drawn toward the center of mass of the object during the collapsing phase, its negative energy counterpart if it was present in the same location at the same moment would be repelled in the exact opposite direction by a force of similar magnitude (to the extent that the average cosmic density of positive energy matter can be neglected). Thus, in such a case, we would need to take into account at least one additional binary degree of freedom associated with the sign of energy of the matter particles present inside the surface, which would also determine their momentum directions.

But this would actually be the simplest case, as more complex momentum states would occur if the matter was not contained within a surface that constitutes a black hole event horizon, because under such conditions not only would the momentum directions of the particles be allowed to vary, but it seems that their magnitudes could also vary significantly. It is important

to understand, however, that the validity of the Bekenstein bound would be preserved even if more information was required to determine the exact microscopic state of matter under those less constraining conditions. This is again because while more information may be required to describe the state of matter when the magnitude of energy and the direction of momentum is not fixed, this information gain would be offset by the decrease in gravitational entropy (the amount of information required to describe the unknown microscopic state of the gravitational field itself) that would result from the lower (nearer to zero) positive and negative densities of matter energy associated with such configurations, or from a mixture of matter of both energy signs (I will explain in section 3.7 why it is, exactly, that a local diminution in the magnitude of matter energy density is associated with a lower measure of missing information concerning the microscopic state of the gravitational field).

Now, it may appear contradictory that under ordinary circumstances, when no macroscopic event horizon is present and the distribution of matter energy is smoother, it is more difficult to tell the energy sign of the particles present within a surface. How could it be more difficult, in effect, to determine the microscopic state of the matter when it seems that you can actually see or directly probe more of the content of the surface? But actually the presence or the absence of an event horizon has nothing to do with the fact that it may be more or less difficult to identify the microscopic state of the degrees of freedom of the matter which is contained within a surface, as this difficulty arises merely from the fact that the number of such degrees of freedom grows very rapidly when the energy density of matter is growing locally. Therefore, it appears that the fact that information is missing from the macroscopic description of a black hole is not consequent to the presence of an event horizon, but is rather attributable to the microscopic nature of the degrees of freedom which encode the information about the state of the matter that is trapped inside the object. What is not known to an observer outside a black hole is not what is inside the black hole, but simply the exact microscopic state of the degrees of freedom associated with the event horizon, from which information about what fell into the black hole could be obtained. Of course actually obtaining this information could only occur at the expense of an even larger increase of entropy in the environment, as would be the case for any system in thermal equilibrium from which we would try to obtain minimally coarse-grained information, but in principle the operation could be performed.

The validity of this viewpoint can perhaps only be appreciated when we recognize that the classical gravitational field, as it is described in a general relativistic context, is merely the smooth and predictable statistical average of what is actually a randomly fluctuating quantum field which in its ultimate form would be mediated by the exchange of elementary particles. Indeed, if there are local variations in the curvature of spacetime above those described by the smooth macroscopic configuration of the gravitational field, then it is only natural to expect that if some property of the field was to be measured in a very precise location this usually unobserved substructure would become apparent and the information associated with it would no longer constitute missing information. It is my belief that the existence of such microscopic degrees of freedom in the gravitational field on a surface is what allows the missing information about the state of matter located inside an event horizon to be obtained under proper conditions.

In any case, what's most significant regarding those situations where the entropy associated with the gravitational field is not maximum is that we are necessarily dealing with transitional states which will, in general, continue to evolve until the configuration described in the preceding paragraphs is reached. Thus, the negative energy matter which may be present inside a positive mass black hole will eventually be expelled from the object, while the positive energy matter will necessarily reach the singularity. By releasing all matter with an energy sign opposite its own, a black hole actually *increases* its total mass and therefore the area of its event horizon and this means that its entropy grows larger in the process. We are therefore in a situation where a black hole containing less matter (but not less mass) can have a larger entropy. This counter-intuitive outcome is allowed because in those situations where matter contributes to diminish rather than increase the gravitational field on a surface (a general surface, not that associated with the event horizon of a black hole) it also contributes to reduce the portion of entropy attributable to the gravitational field on that surface, which apparently contributes more to the total measure of entropy than the matter itself.

It should not come as a surprise therefore that when negative energy matter is released outside the surface of a positive mass black hole the total amount of information required to describe both the microscopic state of the matter particles still contained within its surface and their associated gravitational field grows larger. A negative energy particle inside a positive mass black hole does contribute (positively) to the amount of missing in-

formation concerning the microscopic state of the object, but at the same time it reduces the amount of information attributable to the gravitational field, which happens to be larger than that attributable to the matter, so that overall the amount of missing information in the black hole is smaller than it would be without the presence of the negative energy particle. From that viewpoint it certainly appears appropriate that negative energy matter cannot be absorbed by a positive energy black hole, given that this would require entropy to decrease.

The more general situation where only the Bekenstein bound may apply is therefore not incompatible with the results I have derived from a study of stable-state black holes from which all matter with an energy sign opposite that of the object has been expelled. In fact, it seems that there is no real difference between the situation we observe in general when opposite energy particles are necessarily allowed to be present within a surface and that which arises when we are considering the surface delimited by the event horizon of a black hole. Yet the fact that the presence of negative energy matter within a positive energy black hole would only be temporary (even from the viewpoint of an external observer, given that negative energy matter does not experience the metric properties of space and time shared by positive energy observers) and would always give way to a more stable state in which only positive energy matter would remain inside the surface delimited by the event horizon of the object, may suggest that such end states play a role in gravitational physics which is analogous to that which is played by thermal equilibrium states in statistical mechanics. But the real question regarding the Bekenstein bound is how it can be that under the more general conditions in which it applies, the energy and the momentum states of matter particles located within a surface are allowed to vary in a continuous way, in both magnitude and direction, while the measure of missing information encoded on the surface must still be provided in binary form.

What my investigations have led me to understand is that in fact this freedom is only apparent. It turns out that even under the more general circumstances discussed here the magnitudes of the energies and the directions of the momenta of elementary particles are restricted to binary values. What allows me to draw such a bold conclusion is that I have recognized the consequences of the fact that event horizons are actually always present on the shortest distances, where quantum fluctuations in the energy of the gravitational field continuously give rise to the formation of ephemeral Planck mass black holes. It is clear that the fluctuations in energy occurring at the

Planck scale do not all by themselves imply that the energy of particles must be fixed to some maximum value, but the fact that such fluctuations are omnipresent when we reach this scale means that elementary black holes are actually the substance of physical space and time at this level of precision of measurement and if that is the case then it means that matter is always shrouded in the event horizons of those microscopic black holes and therefore we can only conclude that locally it is submitted to the same constraints that would apply in the presence of a macroscopic black hole.

Thus, the energies that could be measured locally would always be of the order of the Planck energy, because the particles trapped within those microscopic black holes would be accelerated to arbitrarily high energies by the gravitational fields present on their surfaces. Indeed, the surface gravitational fields produced by black holes with such small masses would be extremely large, therefore compensating for the short time intervals during which they would actually be allowed to accelerate the particles which are submitted to their influence. It must be clear, however, that there can still occur variations of energy in units smaller than the Planck energy on larger scales, where only average values of the energy of matter and its associated gravitational field are significant and most contributions can be expected to cancel out. The Planck energy must not, therefore, be conceived as a minimum unit of energy (in a more general context), because to the contrary it constitutes a maximum level of energy which must nevertheless be the only possible measure of energy *magnitude* concerning the state of matter at the fundamental level of precision of space and time intervals set by current quantum gravitational theories.

The case of momentum direction is a little more complex, because we are here dealing with a scale at which quantum indefiniteness in position cannot be ignored. This is reflected in the fact that the same elementary unit of surface would actually correspond to every possible direction normal to the surface of an elementary black hole. But even if it may never be possible to associate a classically well-defined direction to the momentum of a particle submitted to the gravitational field of such a microscopic black hole, it remains that quantum mechanically there would exist a definite (but superposed) state of momentum even for particles in such a situation and this state would still be constrained by the configuration of the local gravitational field. In other words, there would still be a constraint on momentum direction to be fixed by the direction of the gravitational field. Thus, I believe that when we are considering the states of particles on the scale of an elementary

unit of volume, corresponding to an elementary unit of area (equal to four Planck areas), the momentum direction of a particle may still vary only in a discrete way, even when no macroscopic event horizon is apparent on a larger scale.

Indeed, as a result of quantum indeterminacy, it is not possible to specify the direction of the local momenta any more precisely than there are elementary surface elements associated with the microscopic black hole in which a particle is trapped. So, each elementary unit of area on the event horizon of a local microscopic black hole still contains the same amount of information as would an elementary unit of area associated with a macroscopic black hole. This is true even if it would be possible to define the orientation of the elementary units of microscopic black hole surface in a very large number of ways, because the momentum direction itself is not determined to any better precision. The orientation of the elementary surface elements of the microscopic black hole could vary in a near continuous way, but given that the momenta of the particles constrained by the event horizon of this black hole are in a state of quantum superposition, then their directions cannot be identified any more precisely than by specifying the value of a discrete degree of freedom associated with a particular one of the surface elements, regardless of the exact orientation of those units of area. Thus, on a local scale there would be a finite number of possibilities (associated with the finite number of surface elements on a microscopic black hole event horizon) for the momentum direction of a particle, which can therefore be specified exactly using a minimum number of binary units of information.

Now, given that there appears to exist a correspondence between the state of a matter particle reaching a black hole singularity (conceived as being merely a maximum density state with finite volume) and a given precise elementary unit of surface on the event horizon of the object, then it would seem appropriate to consider that a precise unit of area on a macroscopic surface that is not an event horizon should in general also correspond with the state of a specific matter particle inside that surface. In such a context it should be possible to associate the missing information which would allow to identify the direction of the momentum of a particle contained in a microscopic black hole present inside a macroscopic surface with some precise element (or perhaps with a precise group of elements) on that surface. Thus, if all the matter particles present inside some surface can be considered to be locally constrained by a microscopic event horizon, then even in the absence of a macroscopic event horizon we would be allowed to assume that the miss-

ing information about the exact state of those particles must be provided in a binary form corresponding to specific elements on the surface enclosing the volume in which the particles are located. But this actually occurs only when we assume that event horizons must always be present locally at the Planck scale, so as to constrain the magnitude of the energies and the direction of the momenta of matter particles on such a scale.

Of course under such conditions more binary units of information would have to be encoded on the macroscopic surface to specify the exact microscopic state of each of the matter particles it contains, because in addition to specifying the handedness of a particle we would now need to determine its energy sign and the direction of its momentum, which is dependent on the sign of energy of the microscopic black hole which is constraining its motion locally[5]. Therefore, the amount of missing information associated with the microscopic state of *matter* inside an ordinary surface would be larger than it would be if this surface was the event horizon of a black hole. In fact, the configurations for which the entropy associated only with the signs of energy and the directions of momentum of elementary particles would be the highest are those where macroscopic gravitational fields would be absent and the areas of the local event horizons associated with the presence of microscopic black holes would be the smallest and would be found in the largest number. But given that this occurs when the total energy density of positive and negative energy matter is the smallest and the distribution of matter energy is as smooth as it can be, then it follows that there would be a compensation between the increase in the amount of missing information required to specify the energy signs and the momentum directions of matter particles and the decrease of entropy related to the reduction in the strength of the gravitational field present on the boundary of the region considered, which would still allow the Bekenstein bound to actually apply.

If this account of the physical degrees of freedom of matter associated with the missing information encoded in the microscopic configuration of the gravitational field on a surface is accurate it means that we would not be justified

---

[5]In the present context it is important to understand that the fluctuations which are responsible for the presence of elementary black holes on the quantum gravitational scale are not dependent on the presence of a matter particle and therefore the gravitational field associated with such a fluctuation is not that of any particle submitted to it, which means that the black hole can actually have a mass sign opposite that of a particle which is under its influence locally.

to assume that there is no longer anything physically significant going on *at* the Planck scale, because in fact the state of matter associated with an elementary unit of area on the surface of a macroscopic black hole would actually also characterize the physical reality which exists when we reach the shortest intervals of space and time. I was able to draw this conclusion only at a relatively late stage of my research program, because for a long period I had assumed without much thinking that the possibility that matter could exist in a negative energy state would imply a cancellation of all quantum fluctuations in energy at the Planck scale, which would not allow for the presence of microscopic black holes on such a scale. But in fact all that is truly implied by the possibility that negative energy states can be occupied is that the fluctuations in energy can occur in both positive and negative territory. Thus, not only do fluctuations associated with positive and negative energy states not compensate one another out at the smallest physically significant scale of space and time, but it seems that their basic distinction actually provides one of the only significant degrees of freedom characterizing the state of matter on such a scale.

The fact that the proposed description of the constraints applying on states of matter trapped by the gravitational field of stable-state black holes can be generalized, in the particular manner described above, to situations in which the density of matter is lower and more homogeneously distributed and particles of opposite energy signs can be present together within a surface strengthens the argument for the existence of a correspondence between certain properties of black holes and general features of the physical systems described by conventional statistical mechanics (the discussion featuring in the following section will add weight to this conclusion). Indeed, I have already pointed out that the situation we have in the presence of a macroscopic black hole containing only matter with one energy sign is analogous, from the viewpoint of gravitational entropy, to a state of thermal equilibrium such as we might encounter in the context of conventional statistical mechanics. But if we are justified to assume that the proposed description of the microscopic degrees of freedom characterizing stable-state black holes can be generalized by assuming the existence of states (the elementary black holes) which are similar, locally, to those equivalent thermodynamic equilibrium states then the analogy could be carried over to the field of non-equilibrium thermodynamics. This is because, in effect, the basic assumption of the thermodynamic theory of irreversible processes is that even systems evolving irreversibly are to be conceived as being locally in a state of near thermal

equilibrium. What we have then is an ensemble of subsystems in a momentary state of near equilibrium exchanging energy and evolving in such a way that static equilibrium is not required at the level of the system as a whole (which in the current analogy would be any matter-enclosing surface) like it would be in equilibrium thermodynamics.

It is true that in the present case the stability of the configurations occurring on the shortest scale would be limited because microscopic black holes are continuously being created and evaporated, but then the local subsystems in the theory of near-equilibrium thermodynamics are also not in states of perfect equilibrium. What is reflected in this particularity is merely the fact that we are here actually dealing with statistical laws applying to randomly fluctuating systems for which deviations away from thermal equilibrium continuously occur locally, even for a system in a state of overall equilibrium. In fact, the situation we would be dealing with in general would be one where a surface may enclose a configuration where a relatively large number of black holes of various sizes and variable stability (including macroscopic black holes) are present and interact with one another. In this context the states of matter particles would be locally constrained, but in a more or less stable way depending on the scale of the phenomena being considered, as in the local subsystems of the theory of non-equilibrium thermodynamics, while the system as a whole would be allowed to evolve irreversibly through the merger of smaller mass black holes into ever more massive ones with larger event horizons. One could hardly think of a more perfect analogy between two theories and I believe that this is not a coincidence, but rather a clear indication that the proposed application of the insights derived while studying the problem of discrete symmetries in the context of the existence of negative energy matter allows a better understanding of the problem of black hole entropy as a pure thermodynamic phenomenon in a quantum gravitational regime. It is clear to me that whatever explanation of the discrete nature of the microscopic degrees of freedom of matter particles constrained by the gravitational field of a black hole would be more accurate than the one provided above would have to derive from a more detailed knowledge of quantum gravitation than is currently available.

2.12   Negative temperatures

It is not a widely known fact that while temperatures are usually confined to positive values, it is nevertheless unavoidable that some physical systems be attributed negative temperatures under certain conditions. Those who have considered the issue have recognized in effect that negative measures of temperature must necessarily occur when we are dealing with certain macroscopic systems characterized by a finite number of energy levels. What happens is that as temperature rises it must in general be assumed that more energy states become available for the constituent particles, so that the amount of missing information or entropy is itself rising. Therefore, entropy must be assumed to be minimum when a system is at zero temperature. But for systems with a finite number of energy levels it turns out that as temperature increases we may reach the point where entropy is maximum and temperature therefore must be considered infinite. This may occur for example in the case of a spin system in a magnetic field where the number of levels of orientation of each nuclei is finite. For such a system the lowest energy configuration is that where all the spins are in the direction of the magnetic field, while the highest energy configuration is that which would occur when all the spins would be oriented in the direction opposite that of the magnetic field. At infinite temperature all spins would be oriented in the most random way, with as many spins oriented in the direction of the magnetic field as there would be in the opposite direction. If we add more energy to a system in such a state we would witness a decrease of its entropy, as more spins would become oriented in the direction opposite the magnetic field and less information would be required to describe the unknown microscopic state of the system.

Given that temperature merely defines the relationship which exists between energy and entropy, if an increase of energy produces a decrease of entropy then it must necessarily be assumed that the temperature has become negative. But if adding more energy decreases the entropy only slightly when it reaches its maximum point at which the temperature is infinite then it means that the temperature is not 'minus zero' but actually 'minus infinity'. Thus, as even more energy is added to the system the entropy would gradually decrease back to a minimum at which point the negative temperature would actually reach the zero value again. In the case of the spin system this point would be reached when *all* the spins would be oriented in the direction opposite that of the magnetic field and no further change could

occur. I may also mention that it was found that when we combine two such systems which happen to have opposite temperatures of equal magnitude the outcome must be a system with infinite temperature. It must be understood that despite common expectation to the effect that temperature is a positive definite quantity, the conclusion that negative temperatures may occur in nature is not just a consequence of adopting some particular definition for what temperature should be or of choosing a particular reference scale for this quantity. Specialists are unequivocal concerning the fact that negative temperatures cannot be avoided in a general context, because they are associated with actual states of any system with a finite number of energy levels.

Now, what I would like to point out is that if the constraints I unveiled in the previous section concerning the microscopic states of matter in the presence of an event horizon are valid, then it would mean that black holes are somewhat similar, from a thermodynamic viewpoint, to those more conventional systems for which negative temperatures are allowed. Indeed, I have explained that in the presence of event horizons the relevant microscopic states of matter can be specified using only one discrete degree of freedom per particle, so that a certain maximum number of microscopic configurations (similar to the energy levels in the conventional theory) must be assumed to exist for black holes of any mass. In fact, given that the number of microscopic degrees of freedom associated with the matter content a black hole decreases continuously as it loses mass, it appears that the objects become increasingly similar to the above described spin systems as they decay through the process of Hawking radiation. This similarity is all the more appropriate given that it would seem that if a positive energy black hole has a positive value of surface gravitational field, then a negative energy black hole would have a negative value of surface gravitational field and knowing that the surface gravitational field is the quantity which is associated with the temperature of a black hole in the semi-classical theory, I'm led to conclude that this temperature itself needs to be allowed to vary not just in magnitude, but also in sign. Actually, this can be considered an absolute requirement in the context where a negative mass black hole would radiate particles with an energy sign opposite that of the particles radiated by a positive mass black hole, while the same changes to entropy would still be required to take place as a consequence of the decay process. Thus, if negative energy matter exists, it would seem that some black holes could in effect be attributed negative temperatures which would be made conspicuous by

the reversal of their surface gravitational fields.

The correspondence with the above described thermodynamic phenomenon involving spin systems is complete, because as a positive energy black hole evaporates through the emission of thermal radiation and its mass decreases toward zero (in positive territory) its temperature would rise until it becomes infinite when the object reaches the Planck mass at which point if we were to continue to remove energy (by actually adding negative energy) its mass would start to increase into *negative* territory with an initial temperature that would be infinite but also negative and which would decrease (toward zero) as the negative mass of the object increases. Of course the dependence of temperature on total energy is reversed in the case of black holes, given that a larger mass black hole would have a lower temperature, but if we consider only the relationships between thermodynamic properties then the analogy is valid. Also, if we were able to combine a positive energy black hole (to which is associated a positive temperature) with a similar negative energy black hole (to which is associated a negative temperature) then what we would obtain is not a zero temperature object, but an object with a larger and possibly infinite temperature (just like when we combine two opposite temperature systems in the conventional theory), because the mass of the resulting configuration would be smaller and such an object would radiate energy at a higher rate. Of course it may not be possible from a practical viewpoint to combine opposite energy black holes so as to cancel their masses, but mathematically the correspondence between the quantities involved is valid and matches the expectations derived from conventional thermodynamics theory.

The fact that the existence of such a beautifully perfect correspondence between the semi-classical theory of black hole thermodynamics and the classical thermodynamics of systems with a finite number of microscopic levels of energy is allowed to occur under the hypothesis that two signs of mass are relevant for a description of the thermodynamics of black holes constitutes an additional argument for recognizing the legitimacy of this theoretically motivated insight. In fact, I'm surprised that the conclusion drawn by specialists concerning the unavoidable character of the concept of negative temperature was never considered to imply that energy itself should be allowed to vary in sign rather than only in magnitude. But as I have always believed that the inherited motivation behind the widespread idea that energy can only be positive originates from the thermodynamic conception of energy as a measure of heat (which is itself a positive definite quantity from a classical

viewpoint), I was quite satisfied when I learned that this most thermodynamic concept of all, the temperature, must itself vary in sign. If there is no reason to assume that negative temperatures cannot have a clear significance in physical theory and if it turns out that they must ultimately be associated with the state of objects whose energy is predominantly negative, then we have one less argument for assuming that the concept of negative energy itself cannot be given clear meaning.

2.13   Summary

Once again I would like to conclude the current chapter by providing a summary of the decisive results which were obtained concerning the various problems which can be addressed in the context of the alternative approach to time reversal that was developed in this chapter. The reader who may want to skip this section can do so without missing any essential development necessary to understand other portions of the present report. The key results are thus the following.

1. It would violate the requirement of relational definition of physical quantities to consider a reversal of the directions of space and time intervals, or those of momentum and angular momentum, or of the sign of energy, or that of charge that does not occur relatively to some remaining unchanged parameter of the same kind and therefore such changes must be considered impossible.

2. The reversal of space intervals produced by the P symmetry occurs relative to the unchanged direction of time intervals and therefore a violation of this symmetry does not imply that the universe is fundamentally lopsided, because this violation of symmetry can be compensated by an appropriate reversal of time intervals.

3. For an asymmetry under reversal of some physical parameter to exist all that is required is that the relevant properties be asymmetric with respect to something.

4. Only a combination of discrete symmetry operations that reverses all fundamental physical parameters and leaves absolutely nothing unchanged can be categorized as inviolable.

5. The notion that absolute directionality should not be allowed cannot be considered to restrict the violation of the P, T, or C symmetry operations, but merely the violation of the combined PTC operation.

6. A time reversal operation cannot consist merely in a reversal of the motions and rotations of objects taking place in a reverse chronological order, but must allow to establish a distinction between a physical system left unchanged by the operation and one experiencing reversed time intervals.

7. A distinction is to be made between the bidirectional concept of time direction associated with the existence of a fundamental time direction degree of freedom characterizing the propagation of elementary particles and the traditional unidirectional concept of time direction associated with changes occurring at the thermodynamic level where the notion of entropy is a meaningful property.

8. The bidirectional or time-symmetric concept of time direction is less restrictive and more distinctive than the unidirectional concept of time direction, because it recognizes the possibility for elementary particles to propagate backward in time and also allows to differentiate between identical particles actually propagating in opposite directions of time.

9. It is the impossibility of actually observing processes from a backward in time perspective that justifies the use of a unidirectional time viewpoint relative to which the physical properties attributed to elementary particles are always those which are apparent from the conventional future direction of time, even when the true direction of time in which the particles propagate is the past.

10. Any time direction-dependent physical property of a backward in time propagating elementary particle which would be positive when considered from the bidirectional time viewpoint would appear to be negative from the unidirectional time viewpoint.

11. Even if momentum is to be left unchanged by a properly defined operation of time reversal it would appear to be reversed along with the space intervals associated with the motion of particles from the unidirectional time viewpoint, because when time intervals are followed

in the wrong direction space intervals are also traversed in the wrong direction.

12. The fact that from the bidirectional time viewpoint charge remains unchanged even as a particle reverses its direction of propagation in time allows this physical property to be used as a means to distinguish time-reversed processes independently from the direction of motion of particles which is necessarily observed from a forward in time perspective.

13. When the time intervals associated with the motion of a particle are reversed as a consequence of applying a T operation this change occurs relative to the unchanged direction of space intervals, so that the same positive space intervals are now traversed in the opposite direction of time.

14. A properly defined operation of reversal of the fundamental time direction parameter cannot give rise to a reversal of the thermodynamic arrow of time given that such a T operation has nothing to do with the perceived direction of motion of particles.

15. It must be required that momentum, as the physical attribute conjugate to space, only reverses when space is reversed, while energy, as the physical attribute conjugate to time, only reverses when time reverses.

16. If the sign of action is to remain unaffected by properly defined P and T symmetry operations, then momentum must necessarily reverse as a consequence of a reversal of space coordinates while energy must necessarily reverse as a consequence of a reversal of the time coordinate.

17. It is necessary to explicitly define space intervals as being reversed by a P operation even though the direction of space intervals is usually assumed to be determined by the direction of momentum, because momentum can be reversed without space intervals being equally reversed when the sign of action is reversed and in such a context it must be recognized that momentum direction is an independent quantity whose specification is not sufficient to determine the sign of space intervals.

18. The time intervals associated with the propagation of elementary particles and the sign of energy must be reversed by T even if traditionally

it is implicitly assumed that both the energy signs and the bidirectional time intervals are unchanged under time reversal despite the reversal of the time coordinate.

19. The spin of elementary particles must remain invariant under a properly defined time reversal operation described from the viewpoint of bidirectional time, even though this physical property would appear to be reversed from a unidirectional time viewpoint relative to which momentum would be reversed while the position of particles would remain unchanged.

20. Charge must be considered to be reversed from a unidirectional time viewpoint under a properly defined time reversal operation T despite what is traditionally assumed, which means that to test the invariance of physical laws under time reversal we need to use antimatter.

21. Under an appropriately defined time reversal operation as experienced from a unidirectional time viewpoint it would be electric fields which would reverse while magnetic fields would remain unchanged and not the opposite, because magnetic fields depend on both the direction of currents and the sign of charge of the source.

22. The charge conjugation symmetry operation C must be understood to consist in a combined space and time reversal operation that leaves the sign of charge invariant from the bidirectional time viewpoint, while it appears to reverse the charge and leave the time intervals, the sign of energy, the space intervals and the momentum unchanged from the viewpoint of unidirectional time, as a consequence of the additional reversal to which those quantities are submitted when time is not followed in the right direction.

23. Despite what is traditionally assumed the direction of spin must reverse from a unidirectional time viewpoint under a properly defined charge conjugation operation given that the space coordinates are reversed while the momentum is left invariant by being reversed twice and in such a context it can no longer be assumed that the behavior of spin under application of C is a mere matter of convention.

24. Handedness must be assumed to be reversed by a properly defined C operation from both the bidirectional and the unidirectional time

viewpoints because from the former viewpoint momentum is reversed and spin is invariant, while from the latter viewpoint momentum is invariant and spin is reversed, which actually explains why particles of a given handedness often seem to be naturally related to antiparticles with opposite handedness.

25. Invariance under a combined PTC operation is explicitly required in the context of the redefined P, T, and C operations which I proposed, because from both a unidirectional and a bidirectional time viewpoint all the fundamental physical parameters are reversed twice or never when the three operations are combined.

26. The classical equations for momentum and angular momentum as a function of space and time intervals and spatial positions do not need to apply from the viewpoint of bidirectional time, because they were formulated in the context of a unidirectional time perspective according to which time intervals are positive definite and it is the space intervals themselves which are reversed. Therefore, it is not possible to argue that the fact that those equations predict outcomes which differ from those provided by the redefined discrete symmetry operations when time intervals are assumed to be reversed is an indication that the new definitions of P, T, and C are inappropriate, because in this context it is rather the classical equations which are inapplicable.

27. There are four different action sign reversing symmetry operations which can be denoted M_I, M_P, M_T, and M_C and whose four different outcomes are each related to phenomenologically distinct states of negative action matter which can be transformed into one another by individually applying the three action sign preserving symmetry operations P, T, and C.

28. There are two different ways by which space- or time-related parameters can be reversed in such a way that the sign of action is reversed, because it is possible to either reverse the signs of the momenta and the energies while keeping space and time intervals unchanged, or else to reverse the space and time intervals associated with the propagation of particles while keeping the signs of the momenta and the energies invariant, but those two different ways to reverse the action can be applied differently to space- and time-related parameters.

29. A negative action particle would propagate negative energies forward in time or positive energies backward in time and would also have negative momentum in the observed direction of its propagation in space.

30. From the bidirectional time viewpoint the sign of charge remains unaffected by a reversal of action, while spin must be assumed to be reversed under all action sign reversing operations.

31. Applying any of M_I, M_P, M_T, or M_C alone once or twice would not necessarily produce invariance, but the $M_I M_P M_T M_C$ operation obtained by combining of all those action sign reversing symmetry operations must necessarily produce invariance given that such an operation reverses all fundamental physical parameters twice.

32. The M_I, M_P, M_T, and M_C operations can be violated to different degrees when applied independently, because the action sign preserving reversal operations P, T, and C which relate the different states of negative energy matter to one another can be violated to different degrees by negative energy matter compared to how they are violated by positive energy matter and it is merely required that the different states of negative energy matter which are related to each other by the action sign preserving symmetry operations be invariant under a combined PTC operation. In such a context the action sign reversing symmetry operations can be conceived as together transforming merely one single additional degree of freedom.

33. Even though I have proposed that it is the existence of negative action matter which is allowing a compensation of positive and negative contributions to vacuum energy density, the fact that we are observing a small positive value for the cosmological constant does not mean that the M symmetry relating positive and negative action states is violated in our universe, because from the viewpoint of the proposed generalized gravitational field equations the imbalance which is responsible for the observed non-vanishing value of the cosmological constant can develop even in the absence of such a violation.

34. When a condition of continuity of the flow of time (associated with the sign of physical time intervals) along an elementary particle world-line is considered to apply it must be considered empirically forbidden for

a given particle propagating a positive charge forward in time to transform into, or to interact with a similar particle propagating a negative charge in the same direction of time, given that the annihilation of an ordinary particle with an ordinary antiparticle must be allowed to occur with the same probability for all pairs and cannot only take place for those pairs where the two particles happen to be propagating in opposite directions of time.

35. If the condition of continuity of the flow of time applies, then no particle can turn into an antiparticle without actually reversing its direction of propagation in time regardless of whether or not it also reverses the sign of its energy. In the context where it would be assumed that all matter must be created out of nothing at the Big Bang this would mean that there should be as many forward in time propagating particles as backward in time propagating particles in the universe, which allows to conclude that no fundamental asymmetry under reversal of the direction of time can be related to the thermodynamic arrow of time. This conclusion is not ruled out by observations given that the most abundant form of negative action matter can consist of backward in time propagating particles. In the context where the condition of continuity of the flow of time must apply, the compensation of the observed matter-antimatter asymmetry which is made possible by the presence of negative action matter is no longer a mere possibility and there must necessarily be an equal number of particles and antiparticles of all action signs taken together, which in fact also means that there must be as many positive action particles as there are negative action particles of any kind in our universe.

36. If we recognize the necessity for a compensation of the positive action matter-antimatter asymmetry by an opposite asymmetry involving negative action matter and antimatter, then in the absence of any preexisting matter in the initial Big Bang state it follows that it must definitely be possible, under the conditions existing in the very early universe, for pairs of opposite action particles to be permanently created out of the vacuum even if this is forbidden under ordinary circumstances, but only if we require the condition of continuity of the flow of time along an elementary particle world-line to actually apply even under such extreme conditions.

37. In the context where the limitations imposed by quantum indeterminacy are assumed to imply the existence of a smallest meaningful unit of area if the degrees of freedom on the event horizon of a black hole are to be associated with the state of some quantum particles associated with the gravitational field, then given that it would be impossible for two particles to go through such a unit of surface at the same moment, it follows that no physical parameter associated with such a unit of area can be attributed more than one value at any particular time.

38. The microscopic degrees of freedom of the gravitational field on a surface must be considered to reflect the microscopic state of the matter that is located within that surface, particularly when this surface is the event horizon of a black hole, even though the degrees of freedom of the matter itself may not be of the same nature as those associated with the surface.

39. An elementary black hole with a mass equal to one Planck mass and an area that is four times that corresponding to a sphere with a radius equal to the Planck length and which we must assume to contain at most one elementary particle should carry exactly one binary unit of information which means that it is possible and even necessary to associate each unit of information encoded on an event horizon with the state of one of the particles it contains which can therefore only vary as a binary parameter.

40. If all distinct degrees of freedom associated with the discrete symmetry operations and only those degrees of freedom needed to be reflected in the microscopic state of a particle confined by the event horizon of an elementary black hole we would need three bits to be encoded on the event horizon of the object.

41. When we restrict our attention to stable-state black holes it must be assumed that the sign of mass of the black hole determines the sign of energy of all the matter particles it contains and therefore the degree of freedom associated with the energy sign of particles, which is transformed by the M symmetry, cannot contribute to the entropy of a macroscopic black hole.

42. It is necessary to specify the handedness of particles independently from the other degrees of freedom which are reversed by the P and

T symmetry operations and therefore this parameter can contribute independently to the entropy of a black hole.

43. The sign of charge of the particles forming a black hole is the only physical parameter that is entirely determined by its dependence on the redefined time reversal symmetry operation T as it would be experienced from the unidirectional time viewpoint.

44. It must be assumed that it is merely the momentum direction of the particles forming a black hole which constitutes the degree of freedom that is transformed by the P operation, because while the handedness is reversed by P along with space directions it can also be reversed when the spin reverses and this is allowed to occur independently from a reversal of space-related properties.

45. If negative energy particles were present inside a positive mass black hole they would be rapidly ejected from the object and in the process would acquire a maximum energy and a momentum which would be invariably directed away from the center of the object in the reference system relative to which the black hole is not rotating.

46. A positive energy particle crossing the event horizon of a positive mass black hole would gain a maximum energy and a momentum invariably directed toward the center of the object in the reference system relative to which the black hole is not rotating, while the lateral components of its momentum would become negligible and would merely contribute to the total angular momentum of the object whose motion of rotation is shared by all particles.

47. The maximum energy that is reached by particles accelerated in the gravitational field of a black hole is the Planck energy associated with the smallest physically meaningful measure of area characterizing an elementary black hole.

48. Given that the energy associated with the rest mass of a particle reaching a spacetime singularity or emerging from one with a mass opposite its own would be negligible in comparison with its kinetic energy then it must be assumed that the magnitude of momentum also constitutes a fixed variable under such circumstances.

49. From the viewpoint of an observer outside a black hole the quantum bounce that is predicted to occur by certain quantum theories of gravitation is actually taking place over the entire lifetime of the object and actually allows the information about the matter that fell through the event horizon of the black hole and the energy contained in its singularity to be slowly released, without violating the second law of thermodynamics, as the black hole itself evaporates.

50. From the viewpoint of an external observer the particles with the same energy sign as that of the black hole in which they are trapped would spend most of their time either collapsing, with maximum momenta directed toward the singularity, or expanding, with maximum momenta directed in the exact opposite direction (as would occur after the quantum bounce takes place) and therefore the detailed configuration of the microscopic degrees of freedom on the event horizon of a black hole must be considered to reflect the state of the matter it contains at the time immediately before or immediately after it reaches the singularity.

51. When the sign of energy of all the particles that became trapped by the gravitational field of a black hole is assumed to be determined by the sign of mass of the object, it follows that the sign of the momentum of all those particles just before they reach the singularity or just after the quantum bounce occurs is also fixed by the mass of the black hole, so that this microscopic physical parameter cannot contribute to the entropy of such an object.

52. If the sign of energy of the matter particles forming a black hole was not considered to be a fixed parameter, then the sign of the momentum of those particles would also be a free parameter that could contribute to the information content of the object.

53. Given that the direction of the momentum of all component particles is fixed for a stable-state black hole it follows that the handedness of particles allows to determine one microscopic binary degree of freedom which varies as a function of the direction of spin and which can contribute to the entropy of the object.

54. Given that the traditional formula for black hole entropy was derived from properties of the gravitational field and given that it would not be

appropriate to assume that the gravitational field provides information about the sign of elementary particle charges associated with other interactions, then it would also be incorrect to assume that the sign of charge of the particles that form a black hole can contribute to the measure of entropy determined by the semi-classical theory of black hole thermodynamics.

55. Information about the microscopic configuration of the electric charges present inside some surface must be provided by microscopic degrees of freedom associated with the electromagnetic field on that surface, which would provide an independent contribution to the temperature of a black hole and the same is true for any other charge and its associated field.

56. The only information which must be encoded in the microscopic configuration of the degrees of freedom associated with the surface gravitational field on the event horizon of a stable-state black hole is that which allows to determine the handedness of every particle contained in the object, using one single binary unit of information for every elementary particle, a conclusion which complies with the fact that the microscopic state of the gravitational field of an elementary black hole with a surface of four Planck areas carries one binary unit of missing information or entropy. This result confirms that the only relevant physical parameters for matter that becomes trapped by the gravitational field of a black hole are those which are transformed by the redefined discrete symmetry operations.

57. Despite the fact that for a general surface, in the absence of an event horizon, the direction of momentum as well as the magnitude of energy of the particles inside the surface would be allowed to vary more freely, the limit to entropy imposed by the Bekenstein bound would still apply even if more information would be required to describe the microscopic state of matter contained inside the surface, because the local reduction in matter density that is involved would mean that less information would be required to describe the microscopic configuration of the gravitational field.

58. When the entropy associated with the gravitational field attributable to a positive mass black hole is not maximum as a consequence of the fact

that some negative energy matter is present within the event horizon of the object along with positive energy matter, the situation is unstable and will rapidly evolve to give rise to a stable-state black hole in which negative energy matter is no longer present. This is unavoidable in the context where the presence of negative energy matter within a positive mass black hole actually contributes to decrease the entropy of the object and it suggests that stable-state black holes play a role in gravitational physics which is analogous to that which is played by thermal equilibrium states in statistical mechanics.

59. Despite appearances, even in the more general case where a macroscopic event horizon is absent the energies and the momenta of elementary particles are still restricted to vary as binary parameters locally, because in fact event horizons are always present on the shortest distances as a consequence of quantum fluctuations in the energy of the gravitational field which give rise to ephemeral Planck mass black holes which constrain the motion of particles present on such a scale.

60. Despite quantum indefiniteness there would still be a constraint on the momentum direction of a particle to be fixed as a binary parameter by the direction of the gravitational field in the presence of a microscopic black hole, because it is not possible to specify the direction of momentum any more precisely than there are elementary surface elements associated with the microscopic black hole and there is only a finite number of surface elements.

61. To each elementary unit of area on the event horizon of a local microscopic black hole is associated the same amount of information as is provided by an elementary unit of area associated with a macroscopic black hole.

62. In the context where there appears to exist a correspondence between the state of a matter reaching a black hole singularity and a precise elementary unit of area on the surface of the object it would also seem possible and appropriate to associate the binary information encoded on the event horizon of any microscopic black hole located inside a larger surface with a finite number of elementary units of area on that surface, so that the ensemble of such elements on the macroscopic surface

would provide a binary measure of the amount of missing information concerning all the matter particles contained inside the surface.

63. The amount of information required to describe the state of matter constrained by the local microscopic black holes (arising as a consequence of quantum fluctuations) which are present inside a given surface is larger than that required to describe the state of matter inside a macroscopic black hole, because both the energy sign of the microscopic black holes (from which depend the direction of the momenta of the particles submitted to their gravitational fields) and the energy sign of the particles themselves are free parameters.

64. Given that it is possible to assume that the thermodynamic description of macroscopic stable-state black holes can be generalized by assuming that locally the states of matter particles are constrained by the presence of elementary black holes which are the gravitational equivalent of local states of fluctuating thermodynamic equilibrium, it follows that the analogy between the physics of black holes and the classical theory of equilibrium thermodynamics can be carried over to the field of non-equilibrium thermodynamics.

65. In the context where a negative energy or negative mass black hole must be assumed to have a surface gravitational field opposite that of a similar positive mass black hole, the fact that the temperature of a black hole is proportional to its surface gravitational field implies that a negative mass black hole would have a negative temperature, which is entirely appropriate given that such a black hole would radiate negative energy particles and in the process diminish its entropy.

66. The fact that a negative temperature can be attributed to a negative mass black hole strengthens the case for an exact correspondence between black hole thermodynamics and the classical theory of thermodynamics according to which negative temperatures are unavoidable when a limited number of energy levels are available for a system as its temperature rises.

Chapter 3

Classical Cosmology

3.1 The outstanding problems of cosmology

The situation we face today in the field of theoretical cosmology can be resumed by mentioning two broad categories of problems. The first issue has to do with dark energies in general and the consequences of the existence of invisible forms of matter and energy on the gravitational dynamics of visible matter. One of the main difficulties regarding dark energies has to do with explaining how it is possible for the density of vacuum energy to be as low as one observes it to be, while not being exactly null. Indeed, with the discovery that the expansion of space is accelerating [26, 27] it has become necessary to recognize that some invisible form of positive energy with negative pressure is present in empty space and in the present theoretical context the only plausible explanation we have for this phenomenon is that it is manifestation of zero-point vacuum fluctuations. But such a small value for the cosmological constant is unexpected and therefore one is encouraged in seeking alternative and more exotic interpretation for this dark energy. In the first portion of the present chapter I will explain that it is in fact still possible to assume that dark energy is a manifestation of the non-vanishing value of the cosmological constant which arises from zero-point vacuum fluctuations and I will show that this hypothesis is not invalidated by the otherwise inexplicably small, but non-zero value of this parameter.

Another aspect of the problem of dark energies has to do with the phenomenon of missing mass which arises because it appears that the visible material that is present in galaxies and clusters does not provide enough

gravitational force to explain the motion of the astronomical objects that compose those large scale structures. Here one of the main objectives usually consists in trying to determine the exact nature of the dark matter particles which are assumed to contribute additional gravitational attraction around visible structures in the positive energy matter distribution. Despite all the efforts which were devoted to this task, this is a problem which has remained unsolved. But, as I will soon explain, in light of the developments which were introduced in the first chapter of this report it becomes possible to explain most of the missing mass effects observed around galaxies and clusters as being another, perhaps more unexpected, manifestation of zero-point vacuum fluctuations. However, the presence of inhomogeneities in the invisible negative energy matter distribution can also be expected to contribute to the missing mass effect under particular circumstances and therefore I will also examine the consequences of the presence of such unconventional dark matter on the formation of large scale structures.

The other broad category of issues we are currently dealing with in cosmology could be called the inflation problem. This may sound paradoxical, as inflation presently constitutes a dominant paradigm for theoretical cosmology and is still believed to offer solutions to many serious problems in the field. If I'm allowed to speak about a problem concerning inflation it is because there does exist a series of issues which where most accurately described by the originators of inflation theory and which have long been considered to be appropriately solved by one or another instance of such a model, until it became clear that the theory actually offers so much predictive freedom that it is nearly unfalsifiable. As the following discussion progresses, it will become clear that what made the inflation paradigm so successful is mainly an absence of alternative solution to the various problems it was originally proposed to address. Given that I believe that the most important contribution of the originators of inflation theory was to show that there does remain decisive, unresolved issues in cosmology, which could perhaps be solved using their theory, then I will not refrain from discussing those issues as a genuine category of problem to which new solutions can be proposed, even in the context where we do not reject the basic idea that there may have occurred a short period of exponentially accelerated expansion in the first instants of the Big Bang.

Two different aspects of the inflation problem will be discussed in this chapter. The first aspect has to do mainly with the problem of flatness, or the fact that the present rate of expansion of matter on the cosmological

scale appears to be set to some unnatural value which requires an extremely precise adjustment of parameters in the initial state at the Big Bang. I will explain that in the context of the progress I have achieved while solving the cosmological constant problem, this difficulty occurs merely as a consequence of our failure to appropriately recognize that the principle of relational definition of physical attributes must also apply to the energy of the universe. The other aspect of the inflation problem which I will address is the horizon problem which has to do with the fact that it is not possible to explain the uniformity of the very large scale distribution of matter energy as being a consequence of smoothing processes that would obey the principle of local causality in a universe whose history begins at the Big Bang. Two further issues actually constitute particular manifestations of the horizon problem. They are the smoothness problem and the problem of topological defects. Actually, the smoothness problem would not exist if it was not for the fact that it is usually assumed that a solution to the horizon problem would leave the universe perfectly homogeneous, therefore requiring an independent explanation for the fact that some inhomogeneities nevertheless remained in the primordial matter distribution which gave rise to present-day structures. It will be shown that inflation is not required to solve this problem and perhaps also that which is associated with the rarity of topological defects given that those difficulties arise merely as a consequence of the inappropriateness of inflation theory as a solution to the horizon problem.

The one truly amazing consequence of the particular approach I will follow in dealing with the horizon problem, however, is that it offers a new perspective on another decisive problem which is not always recognized as a problem for cosmology despite the fact that it can be traced back to the particular boundary conditions which were in effect at the Big Bang. This is the problem of the origin of the arrow of time which is probably the most serious difficulty currently facing cosmology. It is merely the fact that the problem is so old and has remained unsolved for so long that explains that it is often not recognized as a problem for cosmology, as if we had long ago given up trying to resolve it. But the developments which have been introduced in the preceding two chapters and those which will be discussed in the second portion of the current one will allow to confirm the cosmological nature of the issue and will culminate in providing the first-ever plausible explanation of how it can be that a fully time-symmetric fundamental theory can conspire to enforce boundary conditions which give rise to irreversible evolution and the second law of thermodynamics.

We therefore have two broad categories of problem in cosmology which are the problem of dark energies and the inflation problem and which each involve several different aspects. I will first discuss the cosmological constant problem along with the problem of missing mass as particular aspects of the problem of dark energies, which will then allow me to approach the problem of structure formation from a new perspective. The progress achieved while solving the cosmological constant problem will then enable me to provide a satisfactory solution to both the flatness problem and the related issue of matter creation as one particular aspect of the inflation problem. Then I will discuss the horizon problem as another aspect of the inflation problem, but while addressing this issue and the related problem of the origin of primordial inhomogeneities I will contribute significant insights into the nature of gravitational entropy that will provide the necessary means to formulate a definitive solution to the problem of the origin of time asymmetry.

3.2   The cosmological constant problem

One of the key parameters of the standard model of cosmology that remains unexplained is certainly that which we call the cosmological constant. If there is often reticence to assume that the cosmological constant is a manifestation of the energy of zero-point vacuum fluctuations it is certainly because the density of energy contained in the vacuum is currently expected to be much larger than the energy density we may associate with the observed value of the cosmological constant. It appears much more natural, therefore, to assume that we are rather dealing with some dark energy of unknown nature whose density could vary with the expansion of space, like that of matter. If dark energy is merely a material substance with negative pressure then it would appear natural to assume that it should now have a density similar to that of matter, while it seems rather unlikely that vacuum energy would simply happen to have nearly the same density as that of matter (visible and dark) given that the density of vacuum energy is usually assumed to be unaffected by expansion. Thus, either dark energy is not vacuum energy, in which case we have no idea what its material nature is, or we restrict ourselves to known phenomena and we recognize that it must be vacuum energy, in which case we need an explanation for the observed similarity between the current value of the energy density of matter and that of vacuum fluctuations, that is, we need to explain how it can be that the vacuum contains so little

energy and yet does not provide a null contribution to the universe's energy budget, as we usually assume should be the case if some symmetry principle is responsible for the fact that this energy is so small compared to the natural value associated with the quantum gravitational scale, which is more than 120 orders of magnitude larger.

I find it significant that the problem associated with the small value of the cosmological constant is usually recognized to be a disagreement between the viewpoint of experimentalists and that of theoreticians, because from that perspective it becomes apparent that resolving the issue will necessarily require reconsidering the validity of certain hypotheses we take for granted in the current theoretical context. First of all, it must be acknowledged that despite the fact that the empirical determination of a positive value for the cosmological constant contributed to reinforce the traditional belief that any energy density that could be associated with this parameter should probably be positive, this restriction would be totally unjustified in the context of the progress achieved in the first chapter of this report. Thus, vacuum energy, in particular, could certainly have been negative and the only thing we can be certain about is that it is the observer independent sum of all positive and negative contributions to vacuum energy density which would have an effect on the expansion rates experienced by positive and negative energy observers, unlike would be the case with a material substance like quintessence with pressure opposite its energy sign, which would only influence the expansion rate measured by a positive energy observer through its positive energy component, as any smooth matter distribution with both a positive and a negative energy component. Therefore, in the context of the developments discussed in section 1.6 it may perhaps look like quintessence has an advantage over vacuum energy as a candidate for dark energy in that it could produce the desired effect even when the symmetry under exchange of positive and negative energy states is considered to apply and the material contains just as much positive energy as it contains negative energy. But I will show that this is not really the case and that the advantage rather goes to vacuum energy for at least originating from known physical principles applying to known forms of matter, or forms of matter whose existence can be deduced from know principles.

There is a certain similarity between the prediction of an arbitrarily large magnitude of energy in zero-point vacuum fluctuations and the old problem of the ultraviolet divergence of black body radiation which was solved by the creation of quantum theory. I believe that the commonly met suggestion

that a cut-off may come about in the calculation of the density of vacuum energy which would be associated with the quantized nature of space at the most elementary level is certainly appropriate, but it is also insufficient to solve the cosmological constant problem. Indeed, such a cut-off would simply decrease the energy contributions from their potentially infinite values to very large values associated with the scale of quantum gravitational phenomena and those various energy contributions would still need to cancel out in order to produce the much smaller observed value. This is precisely the problem we face right now: the required cancellation must occur by chance out of a myriad of potentially enormous, independent contributions to the energy of the vacuum. The hypothesis of a quantization of space (to which would be associated a maximum theoretical value of vacuum energy density) is certainly quite inevitable, especially in the context of the developments introduced in section 2.11 concerning black hole entropy and the relationship between discrete symmetry operations and the microscopic states of the matter that crosses the event horizon of such an object. But even though this assumption appears to be valid it is simply inadequate all by itself to reconcile the theoretically derived and observationally inferred values of vacuum energy density.

In fact, I believe that we have no choice but to assume that some symmetry principle must be responsible for the almost perfect cancellation that gives rise to the observed small value of vacuum energy density, because under current assumptions there would be virtually no limit to the expected value of this parameter which would then be more likely to have a relatively high positive or negative value. However, I also share Feynman's opinion that it may not be quantum field theory or the preferred grand unified theory which needs to be modified in order to accommodate such a requirement, but rather our current theory of gravitation. Indeed, the generalized gravitation theory I have introduced in chapter 1 has allowed me to identify a new category of matter particles with negative action sign with which we may naturally expect to be associated a contribution to the energy of zero-point vacuum fluctuations which would be opposite that associated with positive action matter particles.

It is true that there are already both positive and negative contributions to the energy of the vacuum in the context of traditional theories, but it is simply too unlikely that the required outcome could arise by chance from an extremely precise cancellation of the countless, independently varying, positive and negative contributions which are normally taken into account. What

I'm suggesting is that there exists a whole new class of contributions whose total energy must necessarily compensate the sum of all currently considered contributions to the energy of the vacuum. Indeed, in the context where there must be a symmetry under exchange of positive and negative energy states, we are allowed to expect that the energy of the vacuum should actually be null, because negative energy observers would necessarily experience vacuum fluctuation processes which contribute energies that are the exact opposite of those contributed by the vacuum fluctuation processes which are experienced by positive energy observers and which are the only type of vacuum fluctuations currently taken into account by conventional quantum field theory. This is a consequence of the fact that while only one category of positive and negative energy fluctuations directly interacts with positive energy matter, both categories of contributions exert a *gravitational* influence on positive energy matter and must be taken into account in determining the current value of the cosmological constant measured by a positive energy observer.

From my viewpoint, the presently considered negative contributions provided by certain particles present in zero-point vacuum fluctuations would become the positive contributions of those same particles in the negative action sector of quantum field theory (that which describes the processes which directly affect negative energy matter other than through their gravitational influence) and the currently considered positive contributions provided by other particles, also present in zero-point vacuum fluctuations, would become the negative contributions of the same particles in the negative action sector of quantum field theory. This would be true despite the fact that, as I explained in section 2.9, there are actually four distinct action reversal symmetry operations which can be violated in different proportions, because when we are considering all possible processes occurring in the vacuum we are actually dealing with the outcome of all those operations combined and as I explained in the same section there must be invariance under such a combination of all action reversal symmetry operations that relate positive energy matter to negative energy matter.

Thus, all currently considered contributions to the energy density of the vacuum, whether they are positive or negative, must have a counterpart of equal magnitude and opposite sign which guarantees a cancellation of all contributions, regardless of the details of the grand unified theory chosen to describe elementary particles and their interactions. It is not the conclusion that there are no unexpected cancellations among the multiple independent

terms which add up to produce the total energy density of the vacuum which is wrong, but the ignorance of the fact that there is a corresponding set of contributions whose only distinguishing feature is that all of its terms contribute energies which are naturally the opposite of those which are already taken into account, as a consequence of the requirement of symmetry under exchange of positive and negative energy states. It is merely the fact that no fully consistent theory incorporating the concept of negative energy matter had ever been formulated that justified the implicit assumption that no contributions of the kind proposed here needed to be taken into account, because from that perspective the whole idea that virtual processes could take place in the vacuum that would interact merely with negative energy matter appeared irrelevant.

The usual remark to the effect that it is highly unlikely that all contributions to the energy of the vacuum could conspire to produce a vanishing density is justified, but only in the context where the only class of contributions which is recognized to exist is that which is associated with those zero-point fluctuations and virtual particles which exert a direct influence on positive energy matter. However, if we recognize the unavoidable character of the assumption that negative action states are not forbidden, then it would seem that we can now predict a vanishing value for the energy of the vacuum. It is no longer necessary to assume that there occurs a miraculous conspiracy that results in the numerous, currently envisaged, independent contributions to vacuum energy density adding up to produce a number several orders of magnitude smaller than those individual terms. It is also no longer required that the details of some grand unified theory be invoked that would allow to derive the existence of such a precisely adjusted set of independent contributions in order for the right outcome to be derived. We are not really looking for compensations among multiple unconstrained parameters, but for an overall cancellation among two identical sets of parameters whose corresponding elements have equal magnitudes and opposite signs, even on the low energy scale at which the symmetries associated with the unified theory are spontaneously broken. This does not mean that there must be a cancellation of energy fluctuations locally at the Planck scale, however, because as I mentioned in section 2.11 even the sign of energy must be considered a variable parameter on such a scale (in the absence of a macroscopic event horizon to constrain the states of matter particles) and it is merely on the scale at which classical gravitation theory applies that a cancellation of positive and negative contributions is allowed to occur.

What's surprising, therefore, is not that the cosmological constant is so small, but rather that it is in effect not perfectly null. But even if this may not be as serious a problem as that of the discrepancy between current estimates of vacuum energy density and the actual value of this parameter provided by astronomical observations (given that in the present case the amplitude of the required adjustment is much smaller than that which would have to occur in the context of a traditional model), it would not be appropriate to assume that the progress achieved so far constitutes a complete solution to the cosmological constant problem. What I will now explain is that despite the fact that it is natural to expect that there should be a perfect compensation between the currently considered contributions to vacuum energy density and the additional contributions arising from the presence of those virtual particles which directly interact only with negative energy matter, it is nevertheless possible in principle for the cosmological constant to take on arbitrarily large values, even though it does appear that, for some reason, the magnitude of vacuum energy density was negligible compared to the magnitudes of positive and negative matter energy density in the first instants of the Big Bang.

Faced with the dilemma presented here I must acknowledge that I initially tried to explain how it can be that we appear to measure a small but non-vanishing value for the cosmological constant by assuming that in fact the cosmological constant is actually null while the effects we attribute to it are not the manifestation of a non-zero density of vacuum energy, but rather the consequence of the presence of a very large scale inhomogeneity in the invisible negative energy matter distribution. Indeed, as I explained in section 1.8, an overdensity of negative energy matter should produce an outward-directed (repulsive) gravitational force on positive energy matter. Thus, if we happen to be located near the center of such a very large scale overdensity of negative energy matter we should expect to observe a 'local' acceleration of the rate of expansion that would merely be the consequence of the presence of this inhomogeneity in the distribution of negative energy matter. In fact, it was also suggested by others that just the opposite might be occurring and that we may be located inside an *underdensity* in the distribution of invisible *positive* energy dark matter, which would exert a similar outward directed gravitational force on positive energy matter.

But it is precisely here that a problem occurs with my own original hypothesis, because it was later shown that the accelerated expansion which was revealed by observations of high redshift type Ia supernovae is incompatible

with any such explanation of the acceleration of expansion. In fact, in the context where there is a constraint on the amplitude of density fluctuations arising from the uniformity of the cosmic microwave background it appears that there simply could not have existed inhomogeneities of sufficiently large magnitude to provide an alternative explanation of the acceleration of expansion. What's more, if we recognize the observational and theoretical necessity of a critical density of positive energy then we have an additional argument to reject such an explanation for the acceleration of the rate of expansion, because we actually need the additional energy that would be contained in the vacuum in order to reach the critical density which cannot be provided by dark matter alone[1].

It must be acknowledged, therefore, that despite the fact that in the context of the developments proposed in the preceding chapters we may expect the natural value of vacuum energy density to be zero, there must nevertheless exist an imbalance between the positive and negative contributions to vacuum energy density. What must be understood is that this imbalance cannot be attributed to a violation of the symmetry under exchange of positive and negative energy states which is a necessary requirement of the constraint of relational definition of physical properties. At this point it is necessary to recall the definition of the cosmological term that was provided by the generalized gravitational field equations developed in section 1.15. There, I proposed that the value of vacuum energy density associated with the cosmological constant measured by a positive energy observer be defined as the difference between the natural vacuum-stress-energy tensors $\boldsymbol{T}_P^+$ and $\gamma^{-+}\boldsymbol{T}_P^-$ (which denote the maximum positive and negative contributions to vacuum energy density that exert a gravitational influence on positive energy matter) based on the following equation:

$$\boldsymbol{T}_\Lambda^+ = \boldsymbol{T}_P^+ - \gamma^{-+}\boldsymbol{T}_P^- \tag{3.1}$$

From that particular viewpoint it would appear clearly inappropriate to consider the existence of a 'bare' cosmological constant distinct from that which

---

[1]The same argument can also be used to rule out the possibility that dark energy could actually consist of gravitationally repulsive negative energy matter of the traditional kind, which would repel both positive energy matter and negative energy matter itself, because such material would contribute negatively to the energy budget and while it would not form local structures it would interfere with current estimations concerning the initial rate of expansion of matter at the Big Bang (which allow to successfully predict the observed abundance of light chemical elements), when its density would be much larger.

would be associated with the energy contained in zero-point vacuum fluctuations, because the Λ parameter is now explicitly defined as a manifestation of vacuum energy, even in a purely classical context.

Now, what is significant in the above equation is the appearance of the metric conversion factor γ^{-+} in front of the negative contribution to vacuum energy density, which indicates that it is in effect the negative portion of the maximum contribution to the energy of vacuum fluctuations $\boldsymbol{T}_P^-$ that cannot be directly measured by a positive energy observer, while it would be the maximum positive contribution $\boldsymbol{T}_P^+$ that could not be directly measured by a negative energy observer. This is what justifies submitting the maximum negative contribution to the energy of the vacuum to the same metric conversion factor as apply to the measures of negative energy matter density effected by positive energy observers, because in the absence of direct interactions it cannot be assumed that the metric properties of space experienced by this portion of vacuum fluctuations are necessarily the same as those experienced by a positive energy observer. In section 1.15 I mentioned in effect that the presence of the γ^{-+} conversion factor is what allows to establish the quantitative relationship between the metric properties of space experienced by negative energy matter as negative energy observers measure them and those experienced by the same matter as positive energy observers measure them. But if the maximum negative portion of vacuum energy fluctuations is directly experienced only by negative energy observers, then from the viewpoint of positive energy observers the measure of energy density involved must be submitted to the same metric conversion factor as apply to measures of negative energy matter density.

It must be understood, however, that the maximum negative contribution to the energy of the vacuum is not the sum of all negative contributions directly experienced by both positive and negative energy observers, but really the sum of all contributions, positive and negative, experienced by a negative energy observer and which would happen to produce a maximum negative outcome. Now, while the hypothesis that the sum of all contributions to the density of vacuum energy which are experienced by a negative energy observer produces a negative number (while the sum of all such contributions which are directly experienced by a positive energy observer produces a positive number) may appear arbitrary, it is actually unavoidable from an observational viewpoint, as I will explain below. Thus, if the measure of vacuum energy density that is contributed by the maximum negative energy term $\boldsymbol{T}_P^-$ is that which is perceived by a positive energy observer, then it must

be submitted to metric conversion. But even though the necessity of such a mapping is justified by the absence of direct interaction between positive and negative energy matter its legitimacy can only be understood based on considerations of a cosmological nature.

First of all, it must be noted that the magnitude of negative vacuum energy density which would be associated with the natural vacuum-stress-energy tensor $\boldsymbol{T}_P^-$ experienced by an observer made of negative energy matter is an invariant quantity which according to the requirement of symmetry under exchange of positive and negative energy states should be the same as that which is provided by the magnitude of positive vacuum energy density associated with the natural vacuum-stress-energy tensor $\boldsymbol{T}_P^+$ experienced by a positive energy observer. Thus, if, in the context where the cosmological term does not vanish to zero, there must be a difference between the maximum positive and the negative contributions to vacuum energy density measured by a positive energy observer, it can only arise because the metric properties of space that determine the magnitude of the negative energy portion of vacuum fluctuations as they are perceived by such an observer are not the same as those that determine the magnitude of the same negative portion of vacuum energy as they are perceived by a negative energy observer. What I'm suggesting is that this means that the appearance of the metric conversion factors in the definition of the net value of vacuum energy density is a consequence of the fact that the volume of space contained within a given boundary varies depending on whether this volume is measured by a positive or a negative energy observer, so that the same invariant maximum contributions to the density of vacuum energy can provide different contributions for observers of opposite energy signs.

Now, when I introduced the notion of observer dependent gravitational fields, which gives rise to observer dependent metric properties, I emphasized that it must be recognized that there is still a correspondence between the local topology of space associated with positive energy observers and that which is associated with observers of opposite energy sign. Thus, the set of events occurring in spacetime must be the same regardless of the way the metric properties of space are perceived, which also means that every particle that is present inside a surface parametrized using the metric properties of space associated with a negative energy observer must also be present in a corresponding surface parametrized using the metric properties of space associated with a negative energy observer, even when the volume contained inside the surface varies as a function of the sign of energy of the observer. In

such a context even if the ratio of the average densities of positive and negative energy matter would be fixed from the viewpoint of any given observer, the average densities of both positive and negative energy matter could be different for observers with opposite energy signs which do not share the same metric properties. The crucial point here is that those observer dependent metric properties may not only differ locally as a consequence of the presence of variations in the densities of positive and negative matter energy, but may also be different on a cosmological scale as a consequence of a difference in the expansion rates measured by opposite energy observers.

To visualize the nature of the relationships between the measures of energy density perceived by positive and negative energy observers on a cosmological scale it may help to consider the analogy provided by the case of a universe with bi-dimensional space and closed geometry. More specifically, we may imagine two spherical surfaces centered on the same point (in three-dimensional space) which would represent the entire volumes of space experienced by opposite energy observers[2]. It would then be appropriate to assume (for reasons that will be discussed later) that initially, in the first instants of the Big Bang, the two surfaces both have minimum areas which correspond to a state of maximum positive and negative energy densities. Under such conditions the average densities of positive and negative energy matter particles determined using the metric properties of space associated with one of the surfaces would initially be the same as those which are determined using the metric properties of space associated with the other surface. But, as space expands and the two closed surfaces grow in size, the slightest difference in their expansion rates would make their respective areas to differ. Yet, even as those differences would develop, to each position of a particle on the smaller surface would still correspond a unique position on the larger surface associated with observers of opposite energy sign and to each boundary on the smaller surface would correspond one larger boundary on the other surface. In the absence of any *local* variations in the metric properties of space experienced by opposite energy observers the only difference which would characterize the matter distributions observed on the two surfaces would therefore be the difference between their average densities, which would follow from the fact that the same particles occupy spherical

---

[2]It must be clear that the situation described here is only valid as an analogy, because, as I will explain in section 3.5, in a more realist context it is not even possible for space to be closed from both the viewpoint of positive energy observers and that of negative energy observers.

surfaces with different total areas.

Even in the absence of local density variations, therefore, it seems that the metric properties of space could differ for observers of opposite energy signs, as the regions of space delimited by corresponding boundaries (associated with observers of opposite energy signs) could have different volumes depending on the sign of energy of the observer that determines this volume, if the rates of expansion measured by opposite energy observers are themselves allowed to differ (or if there was a difference between the initial values of the scale factor determined by observers with opposite energy signs). This is due to the fact that even though it is possible for the scale factor to differ for opposite energy observers, the same events involving the same particles must exist in the universe independently from the sign of energy of the observer, so that the average densities of positive and negative energy matter measured by a positive energy observer are actually allowed to differ from those measured by a negative energy observer, even when those measures are the same from the viewpoint of both observers initially.

I believe that what is implied by the appearance of the metric conversion factors in the proposed definition of the density of vacuum energy, therefore, is that the invariant maximum positive and negative contributions ($\boldsymbol{T}_P^+$ and $\boldsymbol{T}_P^-$) to the energy density of the vacuum can be made to differ as a consequence of the fact that opposite energy observers do not necessarily share the same metric properties of space, even on the global scale where matter can be expected to be homogeneously distributed. The rule would be that when the scale factor is measured as being proportionately larger by a positive energy observer, the density of the maximum negative contributions to the energy of the vacuum (which cannot be directly measured by such an observer) would be reduced from the viewpoint of such an observer in comparison with the density of the maximum positive contributions to the energy of the vacuum measured by the same observer, so that according to equation (3.1) above the density of vacuum energy would be positive and our positive energy observer would measure a positive cosmological constant. This would be due to the fact that from the viewpoint of an observer that measures a larger volume of space on the cosmological scale those vacuum energy fluctuations whose invariant maximum density can only be measured by an observer of opposite energy sign would appear to take place within a comparatively larger volume and would therefore appear to have a lower density and to provide a smaller contribution than the vacuum energy fluctuations whose invariant maximum density our observer can directly measure.

A definite relationship would therefore exist between the net value of vacuum energy density and the difference between the scale factors determined by observers with opposite energy signs which is made even more significant by the fact that the cosmological constant, as a manifestation of the non-vanishing measure of vacuum energy density, must itself modify the rates of expansion experienced by positive and negative energy observers which determine those scale factors. I'm thus in a position to predict that if the current value of the cosmological constant (or the current average value of the density of vacuum energy) is positive, then it means that from the viewpoint of the metric properties of space associated with positive energy matter the universe must expand at a rate slightly higher than would be measured based on the metric properties of space associated with negative energy matter (if we assume that those expansion rates must have been the same in the initial Big Bang state, as I will propose in section 3.5) and this contributes to further enhance any small difference that may have existed initially between the volumes of space determined by positive energy observers and those determined by negative energy observers.

Indeed, what we measure through observations of supernovae explosions is an acceleration of the rate of expansion that is experienced only by observers with our own sign of energy, while observers with an opposite sign of energy would measure a different variation of the rate of expansion given that the same vacuum energy would exert an opposite gravitational force on negative energy matter. Thus, while a positive cosmological constant would accelerate the expansion of space from the viewpoint of a conventional positive energy observer (due to the larger contribution of its negative pressure), it would actually contribute to decelerate the expansion rate for a negative energy observer (again as a result of its negative pressure), thereby further increasing the divergence between the measures of average matter density associated with observers of opposite energy signs. This means that the cosmological constant must be expected to grow with time, as its current positive value will accelerate the rate of expansion observed by positive energy observers, which will give rise to an even higher positive density of vacuum energy and a larger cosmological constant.

To avoid confusion it must be understood that what allows one to assume that a positive cosmological constant produces an acceleration of the rate of expansion of space that is measured by a positive energy observer, and not merely an acceleration of the rate of expansion of positive energy matter, is the fact that the same metric conversion factor that is involved in deter-

mining the net value of vacuum energy density also affects the measure of density of negative energy matter determined by a positive energy observer, as is made perfectly clear in the formulation of the generalized gravitational field equations introduced in section 1.15. Thus, what we may call the *specific density* of negative energy matter (that which is measured by a negative energy observer) actually becomes larger than the specific density of positive energy matter (that which is measured by a positive energy observer) when the cosmological constant is positive. But the presence of the metric conversion factor in the second term of the decomposed generalized gravitational field equations associated with a positive energy observer produces the same reducing effect on measures of negative energy matter density as applies to the negative portion of the natural vacuum-stress-energy tensors and which gives rise to a net positive value for the energy density of the vacuum.

As a result, despite the fact that the average *specific* density of negative energy matter grows comparatively larger, the average density of negative energy matter which enters the gravitational field equations associated with a positive energy observer remains as similar as it originally was to the specific density of positive energy matter (that which is observed by a positive energy observer). Of course a similar effect will occur for the measures of average positive energy matter density entering the gravitational field equations associated with a negative energy observer, because despite the fact that the average *specific* density of positive energy matter becomes comparatively smaller than that of negative energy matter, the average density of positive energy matter that is physically significant for a negative energy observer would actually grow in comparison with that measured by a positive energy observer, along with the specific density of negative energy matter, as a consequence of the presence in the gravitational field equations of the metric conversion factor associated with a negative energy observer, which must give rise to the same unique cosmological constant (so that it must have the opposite effect as that which arises from the metric conversion factor associated with a positive energy observer).

To return to the analogy of the two embedded bi-dimensional spherical surfaces representing the spatial volumes of a closed universe which are experienced by opposite energy observers, we may determine (through indirect cosmological observations) the average density of negative energy matter on the larger surface associated with positive energy observers in a universe with a positive cosmological constant in order to predict the future evolution of the distribution of negative energy matter. But in doing so we would have

to take into account the fact that the surface on which the negative energy particles evolve has a smaller area, so that the distribution of negative energy matter would appear to be inflated as it is projected on the surface over which positive energy particles evolve. The average density of negative energy matter which would be 'observed' on that surface would therefore be lower than the 'real' density which would be determined based on measures of distances associated with the smaller surface on which negative energy particles evolve. As a consequence, the ratio of the average density of negative energy matter to that of positive energy matter obtained while using the measures of area associated with the larger surface would remain as it was when the two surfaces had equal areas. As a consequence, the ratio of the average density of negative energy matter to that of positive energy matter obtained while using the measures of area associated with the larger surface would remain identical to what it was initially, when the two surfaces had minimum areas. This, I believe, is the true significance of the transformation that is accomplished when one considers the stress-energy tensor of negative energy matter in the form under which it is combined with the appropriate metric conversion factor in the generalized gravitational field equations from section 1.15.

If this interpretation is correct it would mean that the average density of negative energy matter over which are measured the inhomogeneities which may affect the gravitational dynamics of positive energy matter is not the specific density of negative energy matter which is measured by negative energy observers, but a measure of matter density dependent on the metric properties of space specific to positive energy observers and which varies as a function of the rate of expansion measured by such observers. Thus, the variation of the average density of negative energy matter is always assessed by a positive energy observer based on the rate of expansion of space related to his own measures of distance and duration, which on a cosmological scale are influenced only by the average densities of positive energy matter and vacuum energy and the same is true for the density of positive energy matter measured by a negative energy observer. This is why the ratio of the average cosmic densities of positive and negative energy matter must be considered to be an invariant quantity that is not affected by the actual value of the cosmological constant.

There is no a priori motive, therefore, to assume that if matter is expanding at a certain rate from the viewpoint of a positive energy observer, then it should expand at the same rate from the viewpoint of a negative energy ob-

server, as in fact such an observer could even observe space to be contracting (even though during the first instants of the Big Bang the rates of expansion measured by positive and negative energy observers should correspond to a high degree of precision, as I will explain in section 3.5). Thus, even the positive cosmological constant must affect positive and negative energy matter in the same way from the viewpoint of a positive energy observer, because any acceleration or deceleration of the rate of expansion would depend merely on the metric properties of space associated with the gravitational field that this positive energy observer experiences, even though the same density of energy of zero-point vacuum fluctuations would influence the rate of expansion of matter in a different way from the viewpoint of a negative energy observer. On the cosmological scale the rate of expansion does not differ depending on the sign of energy of the expanding matter, but depending on the sign of energy of the observer who measures the expansion.

This does not mean, however, that it is not meaningful and important to determine if the current *specific* density of negative energy matter is in effect larger or smaller than that of positive energy matter, because a larger specific density of negative energy matter would have an effect on the formation of certain large scale structures in the negative energy matter distribution and the presence of inhomogeneities in this matter distribution would also have an effect on positive energy matter. In the context where we may have to assume, for reasons I will discuss in section 3.5, that the average specific densities of positive and negative matter energy were almost equal in the first instants of the Big Bang, the observation that the cosmological constant is positive would therefore constitute a significant result. Indeed, when the ratio of positive to negative energy matter densities is initially so close to unity and the cosmological constant later grows to a larger positive value, it follows that the average density and temperature of matter which are measured by negative energy observers have themselves become larger than the density and the temperature determined by positive energy observers, which means that from the viewpoint of negative energy observers non-gravitational forces may play a greater role in the process of structure formation.

It must be emphasized again that the rule invoked above for justifying that the maximum positive contributions to vacuum energy density are predominant when the scale factor determined by positive energy observers is larger than that which is determined by negative energy observers, simply follows from the fact that in such a case the metric conversion factor associated with the measurements of negative energy matter densities effected by

a positive energy observer transforms the specific density of negative energy matter (measured by a negative energy observer) to a smaller value, while the density of the maximum negative contributions to the energy of vacuum fluctuations must be independent from the sign of energy of the observer, so that when it is submitted to the same metric conversion as apply to negative matter energy it would appear to be reduced in comparison with the density of the maximum positive contributions to the energy of vacuum fluctuations, thereby giving rise to a positive cosmological constant. It should be clear, however, that it is really the *specific* value of negative energy matter density measured by a negative energy observer that is transformed by the metric conversion factor which enters the gravitational field equations associated with a positive energy observer and not the measure of negative stress-energy that is observationally determined by a positive energy observer.

If such a transformation is necessary, it is merely as a consequence of the impossibility to directly compare the average density of matter measured by a negative energy observer with the average matter density measured by a positive energy observer on the cosmological scale, due to the fact that the presence of a smooth distribution of negative energy matter exerts no influence on the gravitational field experienced by a positive energy observer. But this does not mean that we have no reason to expect that the cosmological constant can vary with position, because it remains that the metric conversion factors were defined as locally variable parameters and if that is allowed, then there is no a priori motive to assume that variations of vacuum energy density cannot occur above those directly associated with the presence of matter itself (defined as voids in the homogeneous distribution of vacuum energy).

I initially believed that this possibility could be satisfied once it is understood that, as William Unruh demonstrated [28], the energy of the vacuum is actually dependent on the state of acceleration of an observer, which may justify the locally variable character of the metric conversion factors that provide the measure of vacuum energy density, even in the context where the cosmological constant would have a uniform value throughout space from the viewpoint of inertial observers. This would in effect be allowed in the context where the same acceleration can give rise to opposite measures of vacuum energy density for opposite energy observers (as required if this acceleration is to remain a relationally defined property). But given that the variations involved would need to be independent from any particular choice of reference system, then I have come to realize that such an effect would

not be appropriate as an explanation of the locally variable character of vacuum energy density and therefore a problem would appear to remain. In the following section I will explain what the freedom that is allowed by the above proposed interpretation for the cosmological term to vary as a function of position really means and how it actually becomes an advantage of this particular approach.

In any case, it is now possible to explain what the empirical evidence is that supports the hypothesis that it is the positive portion of the maximum contributions to vacuum energy fluctuations that is directly experienced by positive energy observers (while it is the negative portion that is directly experienced by negative energy observers). First of all, one must recognize that if we were to assume instead that the sum of all contributions to the energy of the vacuum which are directly experienced by a positive energy observer actually produces a maximum *negative* number, then a different form of the generalized gravitational field equations would have to be adopted such that from the viewpoint of a positive energy observer the metric conversion factor would rather apply to the positive portion of the maximum contribution to vacuum energy:

$$\boldsymbol{T}_{\Lambda}^{+} = \gamma^{-+}\boldsymbol{T}_{P}^{+} - \boldsymbol{T}_{P}^{-} \tag{3.2}$$

(this equation is to be compared with equation 3.1 above). The problem which would then emerge is that it is difficult to see how the universe could have evolved in such a way that the scale factor experienced by those observers who measure a lower specific matter density (which would now be negative energy observers) could have become so much larger in comparison with the scale factor experienced by observers with an opposite energy sign that the cosmological constant which results from this divergence could have grown into a positive value that is much larger in magnitude than the specific density of negative energy matter.

Indeed, if we were to adopt this alternative definition of the cosmological term, we should expect that any difference that may develop between the scale factors experienced by opposite energy observers would rather tend to be reduced by the gravitational force attributable to the pressure of this vacuum (unlike is the case when the original form of the generalized gravitational field equations applies), which would leave open the question of how such a relatively large divergence as is revealed by various astronomical observations could have occurred in the first place. The problem of deciding whether the maximum value of the density of vacuum energy fluctuations that interacts

with positive energy matter is positive or negative may then at the present moment remain a purely empirical problem that would be solved in favor of the positive value. In such a context the only theoretical requirements which would apply to the vacuum energy terms that enter the generalized gravitational field equations would be those which impose a maximum magnitude to the positive contribution experienced by a positive energy observer and an exactly opposite value to the contribution experienced by a negative energy observer (as a requirement of exchange symmetry).

It is also important to mention that when it is recognized that all positive contributions to vacuum energy must have a negative counterpart of equal magnitude, the whole notion of false vacuum with a larger than usual energy density becomes somewhat irrelevant, at least from a gravitational viewpoint, given that under such circumstances a non-zero cosmological constant can only arise when there exists a difference between the metric properties of space perceived by observers with opposite energy signs and not as a consequence of the actual nature of the processes taking place in the vacuum. Thus, when we say that a symmetry is broken in a low energy vacuum state, what we should really mean is that the matter particles in this vacuum interact in a way that is different from that by which the same particles interact when they are cooled in a different way in the same vacuum, or by which they interact at higher energies. But that does not mean that the vacuum itself is physically different, in particular with regards to its energy content. Of course, given that I have described matter as being a particular manifestation of vacuum energy, I must recognize that the fact that matter can behave in different ways depending on how a symmetry is broken may nevertheless justify that we refer to the products of such symmetry breakings as consisting of different vacuums. In any case, I think that it would no longer be appropriate to argue that as baryonic matter constitutes only 4 percent of the average positive density of energy, then 96 percent of all matter must be considered of unknown nature, because if dark energy, which comprises about 70 percent of the density of positive energy, really is vacuum energy, then a significant portion of it would consist in the exact same matter particles continuously fluctuating in and out of existence in their virtual form.

Now, if one demands an explanation for the smallness of the cosmological constant in the context of the above description of its origin, one would have to explain why it is that the scale factors and the rates of expansion experienced by observers with opposite energy signs (which we may call the *specific expansion rates* of positive and negative energy matter) were so precisely the

same in the very first instants of the Big Bang that they have remained similar until the present epoch. Indeed, despite the fact that a larger specific rate of expansion of positive energy matter would produce a larger positive cosmological constant which would accelerate the rate of expansion of space measured by a positive energy observer to produce an even larger expansion rate, the cosmological constant and the specific rate of expansion of positive energy matter are still relatively small today. Here it would appear that one may have no choice but to invoke the weak anthropic principle. Indeed, according to Steven Weinberg [29] the observed value of the cosmological constant is so close to the maximum limit imposed by the anthropic principle that it would appear that if it is not much larger this may simply be a consequence of the fact that a larger value would be incompatible with the existence of an observer. What I will explain in section 3.5 is that in the context where we impose a requirement of null energy on the universe as a whole, it becomes possible to assume that it is really anthropic selection which alone requires that the density of vacuum energy be as small as it is currently observed to be.

In any case, the fact that according to empirical data it may appear that the cosmological constant has not changed much during the history of the universe does not affect the validity of the approach defended here, because given the current smallness of the observed value of vacuum energy density, it follows that the rate of change of the cosmological constant, which is determined by the very magnitude of this density of vacuum energy, would have been so small until recently that it would likely have escaped detection (although recent data seem to indicate that the cosmological constant may in effect have grown larger). But it must be clear that if the cosmological 'constant' does grow with time, then despite its small present growth rate it will eventually become arbitrarily large, while the average, specific density of positive energy matter will decrease at an ever faster rate and the average, specific density of negative energy matter will eventually begin to rise and if it was not for the mutual gravitational repulsion to which are submitted the ever more massive astronomical objects of opposite energy signs which are developing in the matter distribution, this density could itself become arbitrarily large. We can therefore expect that the magnitude of the cosmological constant will increase more rapidly in the future, as the acceleration of expansion attributable to vacuum energy becomes more significant in comparison with the rate of deceleration which is fixed by the energy density of both matter and vacuum. Such is the destiny of our universe.

To resume the situation, it transpires that the problem of the cosmological constant was complicated by the fact that it no longer appeared possible to explain its value as being the outcome of a symmetry principle when astronomical observations began to show that it is not exactly zero. This is because any violation of symmetry would likely produce a value of vacuum energy density much closer to the natural scale of energy associated with quantum gravitation. What I have explained is that it is the necessary invariance under exchange of positive and negative energy states (which is justified by the requirement of relational definition of physical attributes discussed in chapters 1 and 2) that allows one to expect a perfect cancellation of all contributions to the density of vacuum energy in the absence of a divergence between the scale factors experienced by opposite energy observers, while it is possible to assume (as I will explain in section 3.5) that it is the weak anthropic principle which alone explains that this divergence was not much larger than it could have been initially, thereby allowing the current value of the cosmological constant to be as small as it is observed to be. I believe that the fact that such a relatively simple and efficient solution to what has been called 'the mother of all physics problem' had never been seriously considered is simply a consequence of the preconceived opinion that negative energy matter cannot exist, which is a consequence of both irrational prejudice and what always appeared to be the insurmountable difficulties preventing a consistent description of gravitationally repulsive matter.

3.3 Missing mass and dark matter

In this section I would like to discuss the impact of the developments introduced in the earlier portions of this report on our understanding of the phenomenon of missing mass[3], which is currently believed to always arise solely from the presence of additional, unseen, but ordinarily gravitating positive energy matter. What will emerge from those considerations is that additional effects, similar to those we would normally attribute to ordinary

---

[3]It must be clear that what I'm referring to here is the general phenomenon that is usually attributed to the presence of dark matter and not that of voids in a matter distribution (even though I will suggest that those two phenomena may sometimes be related) and if I choose this slightly ambiguous and rarely used denomination it is because the problem I'm referring to is more general than the dark matter problem itself, which merely consists in identifying a potential candidate for the weakly interacting massive particles which are usually assumed to explain this missing mass effect.

dark matter, must be taken into account when negative energy matter is present in the universe. But while the existence of negative energy matter may not require a complete rejection of the traditional concept of positive energy dark matter, it also means that it is no longer necessary to assume that such dark matter is responsible for most of the missing mass effect observed at the present epoch around visible positive energy galaxies and clusters. Thus, while I will suggest that based on developments which were introduced in section 2.10 it becomes necessary to recognize the existence of a component of dark, but ordinarily gravitating baryonic matter which could be responsible for a small portion of those missing mass effects, I will also explain that, for the main part, the phenomenon of missing mass appears to merely be a particular effect of the presence of energy attributable to zero-point vacuum fluctuations. Before delving into this important issue, however, I will explore another dimension of the dark matter phenomenon which has been altogether ignored until now and which has to do with the gravitational attraction attributable to the presence of voids in the negative energy matter distribution.

I have already mentioned in section 1.6 that certain forces which could not be distinguished from those traditionally attributed to positive energy dark matter would arise from the presence in our universe of negative energy matter. This is because the presence of an underdensity in a uniform distribution of invisible, gravitationally repulsive, negative energy matter would have the same effect on the surrounding positive energy matter as would the presence of an equivalent amount of ordinary dark matter of the attractive kind. Indeed, on a sufficiently large scale, missing repulsive gravitational forces attributable to the presence of an underdense region in the negative energy matter distribution are equivalent to attractive forces directed toward the same region. If the interaction between positive and negative energy matter is governed by the principles enunciated in section 1.14 it would appear that such a phenomenon should naturally occur (or should have occurred at a certain epoch) around ordinary astronomical objects like positive energy galaxies and clusters, given that such structures would repel negative energy matter and thus create underdensities in this negative energy matter distribution that would enhance the gravitational attraction of the positive energy objects.

It must be clear, though, that it is not possible to conclude that a contribution by negative energy matter underdensities to the missing mass effect around visible structures could make contributions of a distinct nature un-

necessary. The problem we would encounter if we were to completely dispose of other contributions to the missing mass effect is the same one I met when I initially tried to explain the acceleration of the rate of expansion of positive energy matter while assuming that the density of vacuum energy is exactly null, by relying instead on the hypothesis that we are located in a very large scale overdensity of invisible negative energy matter. Indeed, the absence of any additional positive contribution to the energy of matter would be problematic given that the presence of this energy (just like the existence of positive vacuum energy) is needed to bring the density of positive energy to its theoretically and empirically required critical value.

What I came to realize, however, is that in fact it would not even make sense to try to do so, because in order to achieve such an objective we would need to assume the presence of a density of negative energy dark matter which would be much larger than the currently inferred density of ordinary dark matter and in which underdensities of sufficiently high magnitude could exist that would explain all of the missing mass effects presently attributed to positive energy dark matter. If the density of negative energy matter was as small as the observed density of visible positive energy matter, it simply wouldn't be large enough to allow a replication of all the missing mass effects around visible structures which are known to involve equivalent matter densities hundreds of times larger than the average density of ordinary baryonic matter. However, by virtue of the requirement of symmetry under exchange of positive and negative energy matter, if we assume the existence of negative energy dark matter (dark from the viewpoint of both positive energy observers and ordinary negative energy observers), we would also have to assume the existence of *positive* energy dark matter with a similar but opposite average density, (in fact this conclusion would only be valid under the assumption that the density of vacuum energy is negligible in the first instants of the Big Bang, as I will explain in section 3.5) while this is the very hypothesis we might have expected to render unnecessary.

If we just forget trying to do away with additional contributions to the average density of positive matter energy, however, we are led to conclude that the existence of additional contributions to the average density of *negative* matter energy is itself also unavoidable, which does have useful consequences. Of course, such negative energy dark matter would produce additional gravitational attraction on ordinary, negative energy matter orbiting negative energy galaxies and clusters around which it would accumulate. But the presence of underdensities in this negative energy dark matter distribution

can also be expected to have produced additional gravitational attraction on visible structures in the positive energy matter distribution at a certain epoch in the remote past, when the average densities of positive and negative energy matter were much larger.

Again, it must be clear, however, that in the context where the amount of dark matter which is presently inferred to exert an influence on visible, positive energy structures already allows to meet the requirement that there must be a critical density of positive energy (when the positive contribution from vacuum energy is taken into account), then, even aside from the above mentioned difficulties, it is not possible to assume that a significant portion of the missing mass effect observed around present day structures is attributable to the presence of negative energy matter underdensities, because if it was the case then we would no longer be able to assume that the average density of positive energy in our universe is critical, as required from both a theoretical and an observational viewpoint. Indeed, any contributions to the energy budget from inhomogeneities in the negative energy matter distribution would cancel out on the largest scale, if those inhomogeneities developed in an originally smooth distribution of negative energy matter (which I will argue to be a necessary assumption in section 3.9), so that they cannot alone raise the average density of positive energy to its critical value. Thus, we have no choice but to recognize that it is unlikely that the presence of negative energy matter underdensities could contribute significantly to the observed missing mass effect around positive energy objects at the present epoch on all but the largest scales, given that the effect is already known to require the contribution of a density of gravitationally attractive matter energy about as large as that which would bring the total density of positive energy to its critical value.

What is important to understand is that the density of missing mass that can be attributed to the presence of underdensities in the negative energy matter distribution is limited at the present epoch due to the fact that the average cosmic density of negative energy matter is itself finite and relatively small compared to the density of matter inside most visible structures. The presence of negative energy matter underdensities can therefore be expected to have accelerated the process of structure formation in the positive energy matter distribution only at the epoch when the density of matter was still relatively large and homogeneous on the scale of the structures considered. Indeed, any missing mass effect attributable to the presence of underdensities in the negative energy matter distribution can only be concentrated

around positive energy structures when negative energy matter is otherwise smoothly distributed on the scale of the structures considered. Thus, the problem we would face if we wanted to explain the missing mass effect which is currently observed around certain visible large scale structures as being a consequence of the presence of negative energy matter underdensities is that the distribution of negative energy matter, like that of positive energy matter, is no longer homogeneous on the scale at which the phenomenon is occurring, while the average matter density is presently also much smaller than the density of matter inside those visible structures.

Yet the possibility that negative energy matter underdensities could have exerted an influence on the gravitational dynamics of positive energy matter in the early universe is real and certainly not undesirable given that, according to certain accounts, despite all the progress which was achieved in the last decades to model the formation of large scale structures, the currently favored theory of structure formation, involving only positive energy cold dark matter, is still inadequate in certain respects. It is my belief that those difficulties may actually be alleviated by recognizing that there once existed significant contributions to the missing mass effect which arose from the presence of local underdensities in a more uniform distribution of negative energy matter. Indeed, given that the average matter density was larger and the matter distribution smoother when the first galaxies formed, it follows that the gravitational attraction attributable to the presence of negative energy matter underdensities was stronger and more localized early on, so that it must have played an important role (which need not be attributed only to cold dark matter) in the formation of those galaxies.

It is necessary to assume, in effect, that when the distribution of negative energy matter was more uniform, as must have been the case in the primordial universe (for reasons I will explain later in this chapter), negative energy matter was not only found mostly in those locations where positive energy matter was absent, as is the case today, and this means that it was submitted to local gravitational repulsion by positive energy galaxies and clusters, which triggered the formation of underdensities concentrated mostly around visible structures. Of course under such conditions there also arose similar missing mass effects attributable to the presence of local underdensities in the distribution of *positive* energy matter which accelerated the formation of negative energy structures. One interesting outcome of those considerations is that if the processes of structure formation can actually be accelerated by the presence of negative energy matter underdensities in the early universe,

then dark matter may no longer need to be as easily subjected to clumping as it would if it was composed of weakly interacting massive particles of the kind usually considered.

In such a context the conclusion drawn from computer simulations of large scale structure formation, that cold dark matter particles appear to be required to trigger the formation of the observed inhomogeneities in the positive energy matter distribution on smaller scales cannot be assumed to rule out the possibility that the phenomenon behind dark matter may not really involve weakly interacting massive particles, as is normally assumed. Indeed, if a certain portion of the effects currently attributed to cold dark matter can actually be assumed to result from the presence of underdensities in the early distribution of negative energy matter, then the fact that we have no serious candidate for the cold dark matter particles means that we cannot reject the possibility that the missing mass effect may not be attributable to the presence of such particles after all.

This is all the more significant given that it is still the case that current models based on the hypothesis that dark matter is composed of weakly interacting massive particles suffer from a lack of power on larger scales, as witness the problems encountered in trying to explain the observed bulk flows (the motion of entire clusters of galaxies relative to the Hubble expansion) and in reproducing the formation of the largest voids in the galaxy distribution. The presence of some very large scale inhomogeneities in the negative energy matter distribution would naturally allow to provide additional power on such a scale even at the present epoch and may therefore help reproduce some of the effects which would otherwise have to be attributed solely to cold dark matter particles, without much affecting the spectrum of temperature fluctuations in the cosmic microwave background radiation. In the present context the absence of weakly interacting massive particles would not undermine our ability to reproduce the level of structure development that is observed on smaller scales, which would have been appropriately enhanced by the presence of more localized underdensities in the primordial distribution of negative energy matter. Those considerations are particularly relevant in the context where it can be expected that any sufficiently large void that would develop in the negative energy matter distribution would amplify its own growth due to the gravitational repulsion it would exert on the surrounding negative energy matter, so that it appears very likely that negative energy matter underdensities with sufficiently large magnitudes to produce sizable attractive forces on visible matter could be present in our universe on a very

large scale.

Anyhow, if we recognize that the presence of negative energy matter is unavoidable it would follow that, as we approach the center of mass of a sufficiently large structure in the distribution of positive energy matter, an increasingly smaller density of such negative energy matter should be present, because a larger fraction of it would not be able to overcome the repulsive gravitational force exerted by the positive energy matter. We can therefore expect the reduction in the density of negative energy matter that is attributable to the gravitational repulsion exerted by the positive energy structure to grow along with the density of positive energy matter. But clearly, this cannot continue indefinitely, because the average cosmic density of negative energy matter over which the underdensity is measured has a finite magnitude, which is actually much smaller than the density of positive energy matter inside most structures which are currently present on all but the largest scale. When the point is reached at which the underdensity of negative energy matter attributable to the gravitational repulsion of the positive energy matter inside a visible structure corresponds with the magnitude of the average cosmic density of negative energy matter itself, it becomes impossible to further reduce the density of negative energy matter. This marks the limit beyond which the density of missing mass attributable to negative energy matter can no longer grow and actually becomes insignificant in comparison with the growing density of positive energy matter (both visible and dark).

Thus, it can be predicted that past the point at which the gravitational repulsion of a positive energy structure would produce an underdensity of maximum magnitude in the negative energy matter distribution surrounding the structure, the missing mass effects attributable to an absence of negative energy matter would reach a plateau and would only marginally affect the gravitational dynamics of the visible matter inside the structure. If it was not for the finite value of the average density of negative energy matter, the contribution of this matter to the missing mass effect inside a visible structure, such as an early protogalaxy, would keep increasing right up to some arbitrarily large value which would be reached at the center of the structure. It is the fact that there is no similar limit to the magnitude of *overdensities* in the positive energy matter distribution that makes this restriction especially significant. Indeed, given that even in the early universe the density of positive energy matter in a galaxy was already larger than the average cosmic density of negative energy matter (which, as I will explain in

section 3.5, must be assumed to be nearly identical to that of positive energy matter), then it follows that any possible contribution to the missing mass effect from an absence of negative energy matter was only significant in the outer portions of the galaxy, where the average density of positive energy matter was just a small fraction of what it was in the central region. But there can be no doubt that the additional forces which must have existed on such a scale as a result of the presence of underdensities in the negative energy matter distribution are in part responsible for the advanced level of development already achieved at this epoch by the observed structures.

I must mention that even if, according to the approach proposed here, the underdensities in the negative energy matter distribution would develop mostly as a consequence of the gravitational repulsion exerted by aggregates of visible matter, this does not mean that all underdensities in the negative energy matter distribution would always be found to harbor positive energy matter overdensities in their centers. Despite the fact that the development of underdensities would be enhanced by the presence of matter overdensities with an opposite energy sign, it would not be impossible for a void in the negative energy matter distribution to exist without the gravitational repulsion of a positive energy object on a sufficiently large scale, because the presence of such a void would produce gravitational repulsion on the surrounding negative energy matter that could allow the structure to persist all by itself once it is created. In fact, this property may under the right conditions give rise to a self-amplifying process which would allow those voids to reach arbitrarily large proportions similar to those of the largest voids observed in the positive energy matter distribution.

Now, if one recognizes that the presence of negative energy matter underdensities would never allow to explain all of the missing mass effects which are observed around visible positive energy structures at the present epoch, then one must admit that there definitely exist additional contributions of unknown origin to positive matter energy in our universe. Faced with the undeniable evidence that a certain form of dark matter must exist, the normal reaction is to seek to identify a weakly interacting particle, different by necessity from all known particles, that might constitute a viable candidate for this dark matter. But for various reasons, despite the fact that all attempts at detecting and identifying such a particle have failed, it is still believed that dark matter should actually consist of particles that do not interact with ordinary matter *only* through the gravitational interaction. I

CHAPTER 3. CLASSICAL COSMOLOGY

believe that what really motivates this view is the fact that if dark matter interacts with the rest of matter only through the very weak gravitational interaction, then it may in effect become impossible to determine the nature of those dark matter particles by experimental means, which justifies that we concentrate instead on trying to identify a particle that does interact with ordinary matter through one of the other known forces. But what if we could deduce from certain observable properties of ordinary matter that there *must* exist positive energy matter particles which can only interact with ordinary matter through gravitational forces?

At this point you may recall the discussion from section 2.10 concerning the fact that when one recognizes that a condition of continuity of the flow of time along a particle world-line must apply under all conditions, then it is empirically required that any given type of particle (say an electron) that propagates a negative charge forward in time cannot decay into, or interact with a particle of the same kind that would propagate an opposite charge in the opposite direction of time and which would otherwise appear to consist of the exact same kind of particle (it would not merely be an antiparticle of the same kind, but would actually have the same sign of energy and the same sign of charge from the viewpoint of unidirectional time). This is because when the condition of continuity of the flow of time (defined in section 2.10) applies, if certain *ordinary* electrons are allowed to propagate negative charges forward in time while other *ordinary* electrons would propagate positive charges backward in time, then certain electrons could not annihilate with certain positrons (those that would propagate an opposite charge in the same direction of time) with which they would nevertheless be allowed to interact, while it is known experimentally that no such a restriction to electron-positron annihilation exists (all known electrons can annihilate with all known positrons).

But in the preceding chapter I have provided strong arguments to the effect that in the context where an antiparticle must be considered to be an ordinary particle propagating the same charge backward in time with reversed energy, such a condition of continuity of the direction of the flow of time along a particle world-line must indeed be imposed if local causality is to be obeyed. What we do from a conventional viewpoint is that we simply ignore the possibility that an electron, for example, may exist that would propagate a positive charge and a negative energy backward in time by assuming, as a matter of coordinative definition, that a positive action electron always propagates a negative charge forward in time while a positive

action positron always propagates the same negative charge backward in time, as if there were no other possibilities. This is similar to what we do when we exclude negative energy states propagating forward in time, or positive energy states propagating backward in time by assuming that they are unphysical states of matter. In the present case, however, it is not even understood that in doing so we are deliberately choosing to exclude certain states of matter from our description of reality, because it looks like all that is involved is a definition. But that is not the case and if the choice of which positive action electrons propagate a negative charge forward in time and which propagate a positive charge (along with a negative energy) backward in time is in effect a simple matter of definition, the decision to exclude as unphysical those electrons which according to this definition would propagate a positive charge forward or backward in time can only be justified on the basis of observational evidence.

One may argue that this distinction is irrelevant, because the validity of the traditional approach is in fact empirically confirmed, given that it does provide a theoretical framework whose predictions agree perfectly well with observational constraints. Or does it? We still have a serious problem in theoretical cosmology, because we do not know what most of the matter in our universe is made of. Could it be that there is in fact something wrong with some of the implicit choices which were made a long time ago while we were trying to make sense of the newly developed mathematical framework of quantum field theory, before everybody even knew about the existence of dark matter? Is it possible that there does exist in our universe positive action electrons with positive *bidirectional* charges (the measures of charge which are independent from the direction of propagation in time) and positive action protons with negative bidirectional charges and that those particles actually constitute a non-negligible portion of the normally gravitating dark matter, along with the positive action neutrons composed of negatively charged up quarks and positively charged down quarks propagating forward or backward in time?

I do recognize that there may be serious difficulties with this idea, because even if one acknowledges the fact that from an empirical viewpoint positively charged electrons propagating either forward or backward in time should not be allowed to interact with, or to transform into ordinary electrons propagating negative electric charges in any direction of time, or to interact with ordinary protons propagating positive electric charges forward or backward in time, one still needs to explain what justifies this limita-

tion from a theoretical perspective. What's more, even if we could justify the absence of interactions between ordinary matter particles and their dark matter counterparts with reversed bidirectional charges, then it would remain to explain why it is that those particles do interact through the gravitational interaction. I would like to suggest, however, that those difficulties do not decisively rule out the existence of such baryonic dark matter particles and that it is possible to understand, by making use of certain developments introduced in earlier portions of this report, why reversed bidirectional charge particles should in effect be dark, despite the fact that they can also be expected to interact gravitationally with the rest of matter, thereby allowing them to contribute to the missing mass effect around visible positive energy structures.

What I have come to understand is that the difficulty we face while trying to explain the absence of electromagnetic interactions between electrons with negative bidirectional charge and electrons with positive bidirectional charge arises merely because we ignore the fact that the previously defined constraint regarding the continuity of the flow of time along a particle world-line (see section 2.10) must also apply in the case of the particles that mediate the interactions between elementary particles of matter. The problem is that, according to the current interpretation, the world-lines of interaction bosons would appear to abruptly come to an end when they are absorbed by a matter particle, just like they would seem to come into existence discontinuously when they are emitted either by a fermion or another interaction boson. While this may not appear to violate any principle, a certain tension clearly exists between the traditional description of those absorption and emission processes and the previously discussed constraint regarding the continuity of the flow of time along a particle world-line. But, instead of arguing indefinitely as to why such discontinuities are allowed to occur despite the fact that they may be at odds with certain rules that seem to apply in the case of fermions, I would suggest that we simply assume that in fact the flow of time along the world-lines of elementary particles is never really interrupted given that the bosons mediating the interactions between elementary particles of matter somehow allow charges to propagate along two opposite directions of time all at once, *as if* interaction bosons were composite particles made of a fermion and an anti-fermion which need not carry the same charges.

One important characteristic of such an alternative description is that if there must in effect be a continuity of the flow of time along the world-lines of all elementary particles, then in the context where the interaction bosons

would allow a propagation of charges along two opposite directions of time all at once, it follows that the direction of propagation in time of the interacting particles would actually be allowed to remain unchanged during any such interaction process, because time flows in and out of the interaction boson at each vertex. This is of course in accordance with the previously stated conclusion from chapter 2 to the effect that the condition of continuity of the flow of time along the world-lines of elementary particles forbids the transformation of a particle into an antiparticle and therefore it seems that it is really the necessary continuity of the flow of time that imposes that the interaction bosons be described as always propagating charges in two opposite directions of time all at once.

From the viewpoint of this equivalent description of interaction processes, it would follow that for any interaction vertex, time would flow from the incoming fermion into the interaction boson and from the interaction boson into the outgoing fermion (or from the outgoing fermion into the interaction boson and from the boson into the incoming fermion if this particle is propagating backward in time) and the same must be happening at the other vertex of an interaction diagram. An examination of the diagrams describing the interactions between elementary particles, such as those represented in figures 3.1 and 3.2, clearly shows that this hypothesis agrees with the description of all known interaction processes, even those that involve a variation in the charges of the interacting matter particles that must be carried by the interacting bosons, but only when we assume that the bidirectional charge (that which is attributed to known and unknown matter particles which are propagating forward in time when they are observed from the unidirectional time viewpoint) must remain normal (retain the sign of bidirectional charge which is normally attributed to known particles of the kind involved) along the direction of the flow of time associated with the world-lines of elementary particles.

Now, while the above defined condition is satisfied for those processes where the interacting particles are propagating a certain bidirectional charge with a unique given sign either forward or backward in time, it cannot occur for the same processes where only one of the interacting particles is propagating a reversed bidirectional charge (in any direction of time). In the latter case, a bidirectional charge would have to transform, in the direction along which time is flowing, into an opposite bidirectional charge, or a charge propagating either forward or backward in time which is opposite that which is normally attributed (from the viewpoint of unidirectional time) to the parti-
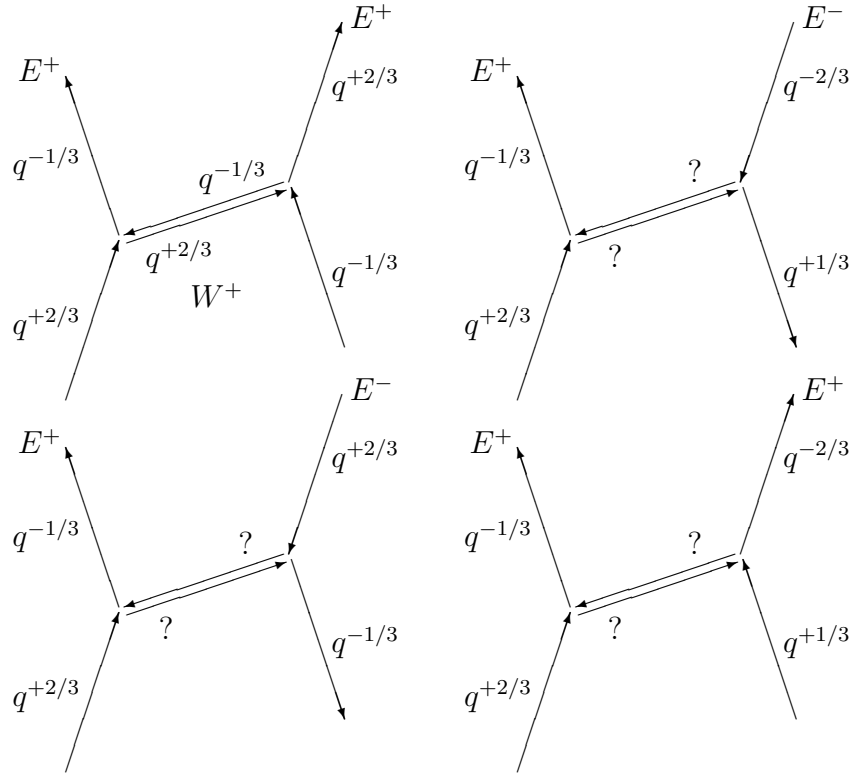
Figure 3.1: Equivalent Feynman diagrams for flavor changing electroweak interactions. Here $q^{+2/3}$ and $q^{-1/3}$ represent the magnitudes and the signs of fractional electric charges as determined from a bidirectional time viewpoint, while E^+ and E^- are the energy signs relative to the direction of propagation in time, which is denoted by the direction of the arrows. The upper left diagram represent processes which are allowed to occur, while the other diagrams represent processes which are not allowed to occur either based on the traditional requirement of conservation of charge or based on the requirement that the normal sign of the bidirectional charge be left invariant along the direction of the flow of time.
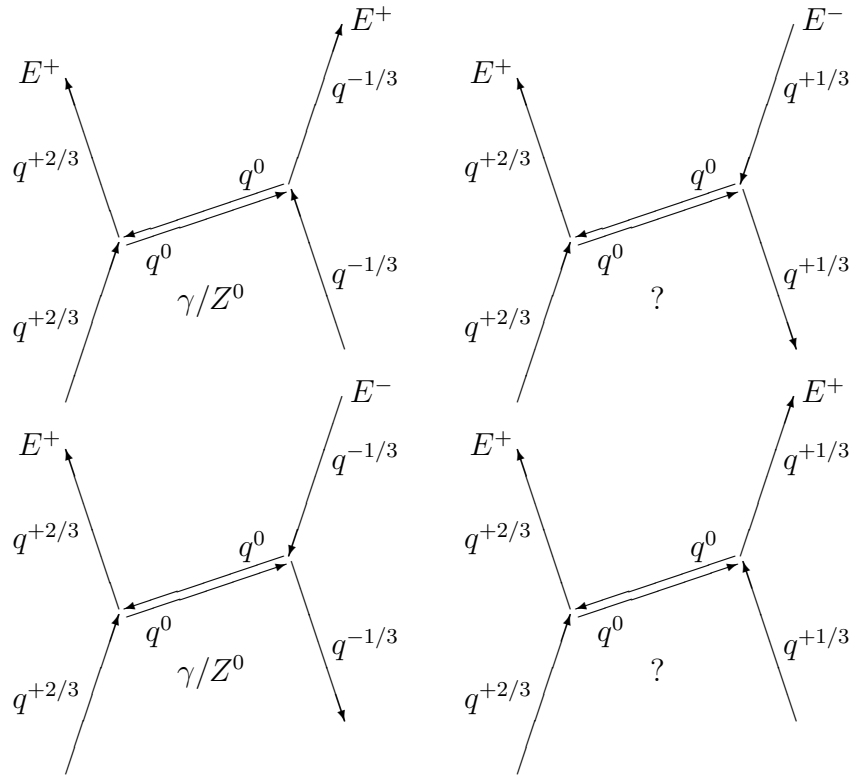
Figure 3.2: Equivalent Feynman diagrams for flavor conserving electroweak interactions between quarks. Here again $q^{+2/3}$ and $q^{-1/3}$ represent the magnitudes and the signs of fractional electric charges as determined from a bidirectional time viewpoint, while E^+ and E^- are the energy signs relative to the direction of propagation in time. It is only for processes of the kind described in the diagrams on the left that the normal sign of the bidirectional charge carried by the interacting matter particles does not vary discontinuously along the direction in which time is flowing and what is observed is that only processes of this kind actually occur in nature, even if from a conventional viewpoint charge would appear to be conserved in all four cases.

cle that carries it when it is propagating in the same direction of time as the particle with which it is interacting (the *normal* sign of bidirectional charge associated with this particle). The crucial point is that even when a neutral interaction boson is involved, the normal sign of bidirectional charge must be carried from the original interacting matter particle into the forward propagating component of the boson and then into the other interacting matter particle, despite the fact that it may appear like no specific normal sign of charge is propagated by the boson.

What I'm suggesting is that it is the fact that a physical attribute of elementary particles associated with the sign of their non-gravitational charges would have to vary discontinuously along the direction of the flow of time that explains that from the viewpoint of the above description of interaction processes the only interactions which are allowed to take place are those involving identical particles (one of which may be an antiparticle) with the same sign of bidirectional charge, or particles which both have a (not necessarily identical) normal sign of bidirectional charge. Thus, it would only be in those cases where the sign of charge remains normal (even when it is actually transformed) along the direction in which time is flowing in the diagram describing an interaction among elementary particles (that which is indicated by the direction of the arrows) that the interaction would actually be allowed to occur, although this is only explicitly apparent in the case of an interaction during which there is an exchange of charge that is carried by the interaction boson.

What's important to understand is that even if in certain cases we would not merely observe a reversal of charge when we follow the direction of the flow of time along a particle world-line, because the particles involved in the process (say an up quark and a down quark) are not of the same type, it is nevertheless required that the sign of charge remains normal along such a particle world-line, even if this may not explicitly appear to constitute a necessary condition when time is actually flowing from a particle with a given type of charge (like a fractional electric charge with a magnitude of 2/3) to a particle with a different type of charge (like a fractional electric charge with a magnitude of 1/3) and the charges of the interacting particles do not vary as a result of the interaction (so that from a traditional viewpoint no charge would appear to be carried by the interaction boson). What I'm proposing, in effect, is that even when two interacting particles do not have the same type of charge and those charges are not altered by the interaction, the sign of charge must still remain normal along the direction in which the particles

involved are propagating in time.

Thus, if an ordinary particle (which is known to interact with other ordinary particles) has a positive charge of $q = +2/3$ when it propagates forward in time, while another such particle has a negative charge of $q = -1/3$ when it propagates forward in time, then it must be assumed that a certain attribute of the particles associated with the normal sign of their bidirectional charges cannot flow continuously from a particle of the first kind with charge of $q = +2/3$, into the interaction boson and then back into a particle of the second kind whose bidirectional charge $q = +1/3$ would be opposite that of a normal instance of such a particle. What allows me to conclude that there exists such an attribute of elementary particles is the fact that there is always a clear (even though relationally defined) distinction between what constitutes a particle and what constitutes an antiparticle, even when we are dealing with particles which do not carry the same types of charge or the same normal bidirectional charge signs (as an up quark and a down quark) and this means that even ordinary particles with opposite normal charge signs must share a certain physical attribute which only varies when the sign of the bidirectional charge carried by those particles (that which is independent of the direction in which a particle is propagating in time) reverses, while it remains unaffected by a mere reversal of the direction of propagation in time that actually leaves the sign of charge invariant. But I must acknowledge that we will probably only be able to fully understand what justifies the rule described here when we obtain a more complete theory of elementary particles which would allow a description of quarks and leptons (and perhaps also of interaction bosons) as composite particles.

In any case if the constraint of continuity of the flow of time extends to interaction bosons in the way suggested here, then it would appear that no interaction can occur that would involve two identical matter particles with opposite bidirectional charges (those observed while following the direction of propagation in time of the particles) propagating in any direction of time, or merely two different particles when only one of them is propagating a charge sign opposite that which is normally propagated forward in time by such a particle in either the past or the future direction of time. I believe that this would be a simple consequence of the fact that no such an interaction could ever be described as a process during which all non-gravitational attributes of the elementary particles involved remain unchanged (are not subject to discontinuous reversal) as we follow the direction of the flow of time along their respective world-lines, from one of the two interacting matter particles

into the interaction boson and then back into the other matter particle, either forward or backward in time.

What must be clear is that there is a difference between the description of an interaction process for which the bidirectional charge sign remains normal along the direction in which time is flowing and the alternative description for which this condition cannot be satisfied. From my viewpoint this distinction is such that it forbids the processes from occurring when the sign of charge would not remain normal, or would reverse along the direction of the flow of time. Thus, even if it may appear that a quantum mechanically equivalent description of a certain interaction process could exist that would be obtained by simply reversing both the sign of charge and the direction of propagation in time for one of the interacting particles, we would have to conclude that the description for which the sign of charge would not remain normal along the direction of the flow of time is actually distinct from that which does not involve such a reversal and may therefore be prevented from occurring. This distinction would simply be a consequence of the fact that, while the sign of charge that is reversed in the apparently equivalent description of the process would not be reversed in relation to the direction of time in which the particle itself is propagating, from a bidirectional viewpoint this charge would nevertheless be reversed. This is what justifies the rule that even when the two interacting particles do not have the same type of charge (as an up quark and a down quark) and are not transformed by the interaction, they still interact only when they both have the signs of charge which such particles normally have (the signs of charge carried by known instances of those particles which are not antiparticles) or when they both have signs of charge opposite those which such particles normally have.

Therefore, I'm allowed to conclude that the rule which is implicitly assumed to apply from a traditional viewpoint, to the effect that no positive action particle that would be propagating charges opposite those of ordinary particles in the opposite direction of time need be considered to exist, is only appropriate in the sense that it is not possible for any such particle to interact with ordinary particles, at least through the exchange of interaction bosons associated with non-gravitational forces. But it is also very clear that this does not mean that particles with reversed bidirectional charges cannot exist, because from a theoretical viewpoint this conclusion would be as unjustified as that which would amount to argue that ordinary particles themselves cannot exist. Indeed, the distinction between electrons propagating positive charges backward in time and ordinary electrons prop-

agating negative charges forward in time is only a relational distinction, in the sense that a positively charged electron propagating backward in time can only be distinguished from an ordinary electron through the fact that it actually has a charge that is opposite that of the ordinary electron, even while it propagates in a direction of time opposite that in which an ordinary electron propagates, but those are not absolutely characterizable properties and an electron with positive bidirectional charge is only different from an electron with negative bidirectional charge in the exact same way an electron with negative bidirectional charge is different from an electron with positive bidirectional charge and it is not possible to distinguish one from the other except through those mutual relationships. If there is no intrinsic or absolute distinction between particles in those two different states, however, then it means that none of them can be considered more real that the other. In other words, both kinds of particles must be assumed to exist, even though matter with reversed bidirectional charges must by necessity be dark from the viewpoint of ordinary matter.

Now, obviously, the only way that such a conclusion could come out as not totally meaningless is if the gravitational interaction is not affected by the condition of continuity of the flow of time along the world-lines of elementary particles, because otherwise there should be no interaction at all between ordinary positive energy particles and positive energy particles with reversed bidirectional charges. But I believe that this is actually unavoidable, because it is clear from the above discussion that it is merely the non-gravitational attributes of elementary particles that must not be subjected to any discontinuous reversal along their respective world-lines. The gravitational interaction is fundamentally distinct from all other interactions in this respect, given that it is neutral with respect to all non-gravitational charges, which is not really the case with other neutral interactions that couple to charge (even though they are mediated by interaction bosons that do not appear to carry a charge). This essential distinction, which is unique to the gravitational interaction, appears to be what allows opposite bidirectional charge particles with the same energy sign to interact gravitationally (and attractively) with one another. The fact that gravitons couple only to energy, while the sign of energy or action is not affected by a reversal of bidirectional charge means that gravitation is the only truly neutral interaction, which therefore remains unaffected by the condition of continuity of the flow of time that prevents the existence of other interactions between opposite bidirectional charge particles.

In any case it seems that the only conclusion that can be drawn is that despite the fact that positive energy matter with reversed bidirectional charges is dark, it would actually exert attractive gravitational forces on ordinary positive energy matter particles, as well as indirect repulsive gravitational forces on all negative energy matter particles, regardless of their bidirectional charge signs. In the first portion of the current section I have explained, in effect, that it is necessary to assume that negative energy (negative action) matter also exists in dark form. But as a matter of principle and due to the requirement of symmetry under exchange of positive and negative energy matter it must be assumed that any such negative energy particle is gravitationally attracted to other negative energy matter particles and is gravitationally repelled by all positive energy matter particles, regardless of their bidirectional charge signs. What allows the existence of repulsive gravitational interactions between negative action particles with reversed bidirectional charges and visible positive action particles is the fact that *all* negative action particles are equivalent to the presence of voids in the positive energy portion of the vacuum, while such voids *necessarily* exert indirect gravitational forces on positive action particles.

Finally, if the conclusion from section 2.10 that there is a compensation of the violation of matter-antimatter asymmetry affecting positive action matter by an opposite violation of symmetry involving negative action matter is correct then we may assume that a similar compensation must apply for positive and negative energy matter with reversed bidirectional charges, given that the same requirement of continuity of the flow of time along a particle world-line that applies in the case of ordinary positive and negative action matter must also apply to matter with reversed bidirectional charges, independently. Thus, the magnitude of the average cosmic density of negative action matter particles with reversed bidirectional charges would need to be exactly the same as the magnitude of the average density of positive action matter particles with similarly reversed bidirectional charges, given that there should be as many particle of one kind as there are of the other kind when there is a condition for all of those particles to be created as pairs out of nothing in the first instants of the Big Bang. This would be a consequence of the fact that the violation of matter-antimatter asymmetry that may explain the existence of baryonic dark matter of the positive energy kind in our universe would have to be compensated by an opposite violation involving baryonic dark matter of the negative energy kind (dark from the viewpoint of both positive energy observers and negative energy observers

made of matter with normal bidirectional charge signs) just as is the case for ordinary positive and negative energy matter. But it is really the magnitude of the average density of positive energy matter with reversed bidirectional charge signs which must be equal to that of negative energy matter with the same bidirectional charge signs, because particles with opposite bidirectional charges cannot be created by pairs (regardless of the action signs of the particles involved). Therefore, the actual density of positive energy matter with reversed bidirectional charges may differ from that of ordinary positive energy matter, as long as there exists a similar imbalance between the average density of ordinary negative energy matter and that of negative energy matter with reversed bidirectional charges.

To summarize what I have discussed so far, it seems that if we are willing to recognize that the existence of negative energy matter is unavoidable, then, in the context where positive energy dark matter must necessarily exist under one form or another, we have no choice but to assume that negative energy dark matter (dark from the viewpoint of both positive energy observers and ordinary negative energy observers) must also be present in our universe if the constraint of symmetry under exchange of positive and negative energy states is to be obeyed. We can therefore expect additional attractive gravitational forces to have been exerted around structures in the visible, positive energy matter distribution in the early universe as a consequence of the presence of underdensities in the distribution of negative energy dark matter, given that variations in matter density were then comparable in magnitude with the average cosmic densities of positive and negative energy matter. As a result, we can also expect the processes of structure formation to have been accelerated by the presence of underdensities in the negative energy matter distribution and under such conditions we are no longer required to assume that all of the missing mass effect must be attributed to the existence of cold dark matter particles.

Now, I must acknowledge that it is not possible to conclude that the missing mass effect is attributable mostly to the presence of particles identical to those that compose visible matter, but which happen to propagate charges opposite those propagated by ordinary matter and antimatter particles in opposite directions of time. Indeed, if most of the dark matter that is assumed to be responsible for the missing mass effect was composed of particles which interact with themselves through the same forces by which ordinary baryonic matter particles interact, then it would be more difficult to

explain the near spherical shape of dark matter halos or certain observations of colliding clusters of galaxies which show that while the detectable high energy gas originally present in the clusters is stripped of the galaxies as a result of such a collision, most of the dark matter is unaffected by the process. I initially thought that this difficulty may simply be a consequence of the fact that we ignore the possibility that due to its higher density, baryonic dark matter would be more susceptible to collapse into stars and other high density objects at a very early stage, so that it would no longer interact with itself on a larger scale when galaxies begin to form later on, which could have allowed to explain the near spherical shape of dark matter halos. Under such conditions it would also have appeared appropriate to assume that the dark matter present inside colliding clusters is mostly unaffected by the collisions, just like the visible stars present in the galaxies, despite the fact that this dark matter is allowed to interact with itself at the particle level.

For this to be a valid hypothesis, however, one would need to assume that a very large amount of positive energy matter exists as massive compact astronomical objects or MACHOs. But even though early studies seemed to indicate that the existence of a large amount of matter in the form of invisible MACHOs was not completely ruled out, because what really motivated the commonly held opinion that there cannot exist enough MACHOs to provide a sizable portion of the dark matter was merely the impossibility for those objects to be formed of ordinary baryonic matter, more recent astronomical observations [30] do confirm that there cannot be a very large portion of ordinarily gravitating matter in the form of MACHOs (regardless of the nature of their constituent particles). Thus, it is no longer possible to assume that a sufficiently large number of such objects could exist that would be composed of baryonic matter with reversed bidirectional charges[4] (which would not have been ruled out by indirect measurements of the density of baryonic matter involving the cosmic microwave background). As a consequence, it is necessary to recognize that the above discussed difficulties associated with the hypothesis that dark matter particles may interact with themselves (like ordinary baryonic matter) can only be surmounted if a large portion of the observed missing mass effect is attributable to a phenomenon distinct from those I have discussed so far. This doesn't mean that none of the dark mat-

---

[4]This is not to say that there cannot exist any large astronomical objects composed of reversed bidirectional charge matter in our region of the universe though, as it is quite possible in fact that small invisible planets made of such matter are present in our own solar system, only this cannot be a very common type of object.

ter can consist of baryonic matter with reversed bidirectional charges (this is not ruled out by the new observations), but merely that it is not possible to conclude that the necessary existence of such matter provides a valid explanation for most of the missing mass effect observed around visible galaxies and clusters.

What I would like to explain now is that it is actually another phenomenon, made unavoidable by the existence of negative energy matter, but not associated with the presence of voids in a matter distribution that is ultimately responsible for most of the missing mass effects. You may recall that I mentioned in section 3.2 that from the viewpoint of the particular interpretation of the metric conversion factors I have proposed and which allows the emergence of a non-zero value for the cosmological constant, it should be possible for vacuum energy density to vary with position in addition to have a non-zero value on the global scale. But in the context of this particular interpretation it appears that if local variations of vacuum energy density do arise, then they could only be attributable to the fact that local differences may develop between the metric properties of space experienced by positive energy observers and those experienced by negative energy observers. What I have come to understand is that in fact such variations are unavoidable, given that the presence of an inhomogeneity in the positive or negative energy matter distribution produces a variation of the metric properties of space which for a positive energy observer is opposite that which is experienced by a negative energy observer.

Indeed, the possibility for opposite energy observers to experience differing metric properties of space as a result of the presence of matter inhomogeneities (which is allowed when it is not possible to directly compare such observer dependent physical attributes) implies that vacuum energy can vary locally, along with the strength of local gravitational fields, as long as there is no compensation between the local gravitational fields attributable to positive energy matter and those attributable to negative energy matter. This is a simple consequence of the fact that different metric properties imply different volumes of space, even locally, and therefore also different measures for the maximum positive and negative contributions to the density of vacuum energy provided by the natural vacuum-stress-energy tensors. It is therefore merely the fact that the variations involved are correlated, under most circumstances, with the presence of local matter inhomogeneities, due to the fact that such variations in the density of matter are usually required to

trigger the development of local variations in the density of vacuum energy, that allows them to provide the long sought explanation of the missing mass effect as being a particular manifestation of dark energy.

What must be understood is that even if *local* fluctuations in the density of negative energy matter can be measured by a positive energy observer and do have an effect on the gravitational field experienced by such an observer, this does not mean that the gravitational fields associated with the presence of matter inhomogeneities of either energy sign cannot give rise to additional effects of a gravitational nature arising from the response of vacuum energy fluctuations to the presence of those gravitational fields. In fact, even in the absence of any inhomogeneity in the negative energy matter distribution, there may arise local variations of vacuum energy density as a result of the presence of positive matter inhomogeneities and the gravitational fields attributable to those local variations of vacuum energy density would actually affect the motion of both positive and negative energy bodies. Therefore, I believe that what explains most of the missing effect around visible positive energy structures is the fact that the gravitational fields produced by those inhomogeneities in the matter distribution give rise to such a local variation of the density of vacuum energy that must necessarily be concentrated around the visible structures and that must give rise to further variations of vacuum energy density arising from the gravitational fields produced by those very same concentrations of vacuum energy.

The problem we would normally face in such a context is that it would seem that the mass of an astronomical object would be allowed to increase without limit as the growth of mass arising from the concentration of vacuum energy would trigger the formation of an even larger concentration of vacuum energy that would further increase the mass of the object. But in fact that is not necessarily a problem, because the energy of the gravitational field generated by a positive energy body is opposite the energy of the source, while the field also interacts with itself, which means that the growth of mass attributable to local variations of vacuum energy should be limited, especially since the gravitational interaction itself is very weak. But this does not mean that no such an effect would exist. In fact, it appears that once one recognizes that negative energy matter itself can exist, one cannot avoid the conclusion that such local variations of vacuum energy density would arise which would have consequences similar to those we normally attribute to the presence of ordinary dark matter, as it would actually contribute to significantly increase the mass of any astronomical object present on a sufficiently large scale.

I must admit that for a long time I, myself, believed that local variations of vacuum energy density could not constitute a solution to the missing mass problem, because I thought that the equivalent mass attributable to such a phenomenon would not be allowed to contribute to the total energy of matter that is required to bring the density of positive energy to its critical value, because there would also be negative contributions to the energy of matter that would arise from those local variations of vacuum energy density attributable to the presence of negative energy matter overdensities, which I thought would cancel out the additional positive contributions, while such contributions also appeared unavoidable. In other words, I had forgotten about the idea, because when I first considered this possibility I thought that given that the energies involved were particular instances of vacuum energy, then both the positive and the negative contributions should add up to produce a null density that would not allow to increase the densities of positive or negative energy to their required critical values.

Also, when I began seriously considering the possibility that some local variations of vacuum energy density attributable to the gravitational field of large astronomical objects could be responsible for the phenomenon of missing mass, I had actually (but inappropriately) come to believe that voids in the negative energy matter distribution could provide an alternative explanation to most of the missing mass effects around visible structures and therefore I didn't see the need that there was to explain the missing mass effect as being the outcome of an inhomogeneous distribution of vacuum energy attributable to the presence of matter, even if the existence of such a phenomenon actually appeared unavoidable. It is only much later that I came to understand that the fact that the distribution of vacuum energy involved would vary with position would make it equivalent, form a gravitational viewpoint, to the presence of local matter inhomogeneities, which may allow one to expect that the negative contributions do not cancel out the positive contributions on the cosmological scale.

It must be emphasized again that what is unique about this interpretation of the inhomogeneous character of the distribution of vacuum energy (which is derived from the generalized gravitational field equations introduced in section 1.15) is that despite the fact that the equivalent mass associated with such a phenomenon actually is a form of vacuum energy that must consequently be dark, it nevertheless contributes to the gravitational dynamics of the universe on a global scale in the same way ordinary matter does. Indeed, if dark matter is attributable to local variations of vacuum energy

density then it must be assimilated with the presence of voids in the otherwise uniform distribution of vacuum energy whenever its energy is opposite the energy of the observer which is experiencing its gravitational field, while a uniform distribution of such underdensities exerts no gravitational force on matter with an opposite energy sign (for reasons which were discussed in section 1.8). Dark matter, therefore, appears to be a hybrid form of matter that shares some properties of vacuum energy or the cosmological constant, but that contributes to local gravitational fields in the same way ordinary matter does, due precisely to its inhomogeneous nature. This means, in particular, that as long as it is uniformly distributed on a global scale, negative energy dark matter, just like ordinary negative energy matter, does not, in fact, affect the rate of expansion of matter determined by positive energy observers and does not contribute to the critical energy density that is relevant to those observers, unlike the negative component of a uniform distribution of vacuum energy and it is only when it becomes concentrated around massive astronomical objects that the presence of this energy becomes apparent to both positive and negative energy observers.

Now, it must be clear that the *average* densities of positive and negative energy dark matter do not change with time, even if the portion of missing mass effects attributable to local variations in the density of vacuum energy only becomes apparent when inhomogeneities develop in the matter distribution and those energies become more concentrated around large astronomical objects. Thus, the additional amount of energy that is present around a positive energy galaxy, but that cannot be accounted for by the presence of ordinary matter, was already present in diffuse form before the formation of that structure, even though it was not exerting any detectable gravitational force *locally* (how this is possible will become clearer once the reader learns about certain unexpected properties of the microscopic structure of gravitational fields in section 3.7). In section 3.5 I will explain that if that was not the case and the average densities of positive and negative dark matter energy attributable to local variations of vacuum energy density were actually growing, contradictions would occur, even if the total energy of matter (comprising the contributions of both positive and negative energy dark matter) was conserved in the process.

From an observational perspective, it would appear possible to confirm that dark matter is a manifestation of spatial variations in the density of vacuum energy, because currently available data indicates [31] that there is a strong correlation, in general, between the gravitational acceleration

attributable to the total amount of matter inside an orbit (say around the center of a galaxy) and the gravitational acceleration attributable to the normal matter. Indeed, if the presence of dark matter must be considered to be an effect of the curvature of space attributable to the matter that is present in a region of space on the local measures of vacuum energy density, then the more gravitational acceleration there is as a consequence of the presence of normal matter, the more distinct the metric properties of space experienced by opposite energy observers must be and therefore the more dark matter there should be. Even though the importance of this empirically determined relationship is often overlooked, it would certainly be a significant problem if it was to remain unexplained, as would be the case from the viewpoint of a more conventional interpretation of the missing mass effect (given that in such a context dark matter is simply an additional component of invisible matter whose existence does not depend directly on the presence of ordinary matter). There is thus a strong motive to prefer an interpretation of the missing mass effect as being a manifestation of local variations in the density of vacuum energy, which must exert gravitational forces similar to those attributable to the matter inhomogeneities that usually generates those variations.

It is important to point out, however, that dark matter, as an effect of spatial variations in the density of vacuum energy, would exert its own gravitational field, which would actually allow it to clump just like conventional dark matter, despite the fact that it really is vacuum energy. This would allow the approach proposed here to reproduce the predictions of the traditional cold dark matter model regarding cosmological evolution and structure formation when an additional contribution to gravitational instability is provided by the presence of voids in the negative energy matter distribution. But it also means that the observations which indicate that large overdensities of visible matter can sometimes become separated from their dark matter component (as a result of collisions between galaxy clusters or in the course of galaxy mergers) can be easily explained, unlike would be the case if the currently unexplained correlations discussed above were the result of a more profound modification of the laws that govern the gravitational dynamics of astronomical objects. Indeed, once created such a dark matter object could continue to exist all by itself for a while, sustained merely by its own gravitational field, just like voids in the matter distribution, while only a minimum measure of dark matter would be left in the visible structure that gave rise

to it[5]. This is a considerable advantage which once again appears to confirm the validity of the generalized gravitational theory developed in the earlier portions of this report.

To conclude this section I would like to briefly return to the problem of black hole information and entropy which was discussed in section 2.11. An important conclusion at which I arrived while trying to determine the nature of the microscopic degrees of freedom of the matter particles captured by the gravitational field of a black hole is that the portion of missing information which is encoded in the microscopic degrees of freedom of the gravitational field on the surface of a stable state black hole would only allow to determine the handedness of each and every matter particle. The other physical parameters characterizing the microscopic state of those matter particles are all fixed to common unique values as a result of the constraints imposed by the gravitational field that is present in the vicinity of the inner singularity. I also explained that, by necessity, the missing information concerning the sign of charge of matter particles (which is transformed by the redefined time reversal symmetry operation T) would need to be encoded in the microscopic state of the field of interaction associated with this charge and is not reflected in the microscopic configuration of a black hole's surface gravitational field. It is only under such conditions that one can obtain the right measure of missing information (that which is determined by the semi-classical theory of black hole thermodynamics) in the case of elementary black holes which contain at most one matter particle.

However, in the context where the sign of charge of a most elementary particle may not only differ as a consequence of a reversal of the direction of propagation in time, but may also be different for particles with opposite bidirectional charges propagating in the same direction of time, one may wonder whether it would still be possible to determine the direction of propagation in time of a given particle from information contained in the microscopic state of the field of interaction associated with the sign of its charge? This is an important question, because if it is not possible to assess the direction of propagation in time of a particle that was captured by the gravitational field

---

[5]From that viewpoint it would appear that the galaxies which appear to contain no dark matter are not galaxies which produce no local variations of vacuum energy density at all, but merely galaxies for which the local variation of vacuum energy has not yet had the time to give rise to additional, observationally significant local variations of vacuum energy density which would themselves produce additional growth of a similar nature.

of a black hole, then one would have to conclude that some physically significant aspect of the state of matter particles cannot be uniquely determined from the information that is contained in the microscopic state of the fields of interaction on the event horizon of such an object, which would imply that information is lost when matter is submitted to gravitational collapse. It may therefore appear that if reversed bidirectional charge particles are allowed to exist there would be a problem with the fact that the units of information concerning the sign of electric charge that would be provided by the microscopic degrees of freedom of the electromagnetic field on the surface of a black hole would not allow to differentiate between a positively charged electron propagating forward in time and an ordinary positron, while there clearly exists a degree of freedom associated with this physical property, which normally allows to differentiate between matter that is visible and matter that is dark.

The above discussion, however, makes it clear that it need not be the case that information about the sign of bidirectional charge is lost, precisely because particles with reversed bidirectional charges would need to be dark from the viewpoint of an observer made of ordinary matter. Indeed, the field that contains the information about the sign of charges or the direction of propagation in time of ordinary matter particles is not the exact same field as that which contains the information regarding the sign of charge or the direction of propagation in time of particles with reversed bidirectional charges. It is the microscopic state of the electromagnetic field with which positive bidirectional charge electrons interact that contains the information about the direction of propagation in time of those particles and given that one can differentiate between this field and that which is produced by electrons with negative bidirectional charges, then it is possible to obtain information about both the sign of charge of elementary particles and their direction of propagation in time from a determination of the microscopic state of all components of the electromagnetic field on the surface containing those particles. Thus, while the distinction between ordinary matter and ordinary antimatter is encoded in the microscopic state of the electromagnetic field that interacts with visible matter, the distinction between baryonic dark matter particles and dark antimatter particles of the same kind is encoded in a different component of the electromagnetic field which is that with which only baryonic dark matter particles interact.

This conclusion, which is dependent on the above proposed interpretation of reversed bidirectional charge particles, is actually much more unavoidable

then one may expect. Indeed, as I mentioned in section 2.11, it seems that on the quantum gravitational scale both the sign of energy of an elementary black hole and that of the particle submitted to its gravitational field could be either positive or negative, because it is only in the case of a macroscopic stable state black hole that the sign of energy of the component particles can be considered to be fixed by the sign of energy of the object. In the case of the elementary black holes produced by local fluctuations in the gravitational field we therefore have two variables which are the sign of energy of the black hole and the sign of energy of the one particle that is submitted to its gravitational field[6]. In such a context it follows that for any black hole, either elementary or macroscopic, once the sign of energy of the object which determines the polarity of its gravitational field is determined there are two possibilities for the momentum direction or the sign of energy of any given matter particle which is under its influence, even if this is only true in the most general case, which is that of a black hole that is not necessarily in a stable state (a situation which is usually ignored when we assume that only positive action particles exist).

Of course, as I already mentioned, the presence in a macroscopic black hole of particles with an energy sign opposite that of the object would contribute to reduce the strength of its gravitational field, which means that it would reduce the amount of information required to describe the microscopic state of this gravitational field. Thus, while additional information would be required in such a case to specify the momentum direction of a given particle (which is dependent on the sign of its energy), this would not contribute to increase the entropy of the black hole. But this does not mean that the sign of energy of the component particles is not reflected at all in the microscopic degrees of freedom on the surface of a black hole, because in the more general case of an elementary black hole it clearly must. In any case, once the polarity of the local gravitational field attributable to a elementary black hole is considered to be determined, not only would the handedness of the particle under its influence be allowed to vary, but also would the direction of its momentum, even if the sign of its energy does not contribute to alter the polarity of the gravitational field.

Information, therefore, must be encoded on the event horizon of an ele-

---

[6]It must be clear that the idea that event horizons are constraining the motion of matter on such a scale is only valid as a semi-classical representation and that it is the quantized nature of space and time itself which implies the existence of discrete degrees of freedom for the matter particles present on the quantum gravitational scale.

mentary black hole with a given energy sign that would allow to determine the momentum direction and therefore the sign of energy of the particle under its influence. In fact, if a quantum theory of gravitation is to eventually constitute a unified theory of all interactions it can be expected that additional information would need to be encoded on the smallest physically meaningful surface about the direction of propagation in time of particles with a given action sign, which is the parameter that determines the observed sign of charge of a particle with a given sign of bidirectional charge. But, additional information would also be needed to specify the sign of the bidirectional charge of a most elementary particle that determines which type of non-gravitational field it interacts with, that is to say, information would need to be available to determine whether such a particle is visible or not from the viewpoint of a given observer. Now, what I'm suggesting is that, not only are those the only fundamental parameters which can vary in a discrete way under such conditions and which actually allow to characterize the state of any matter particle on the most fundamental scale, and not only is it possible for the information that is required to determine the value of each of those parameters to be encoded on the surface of an elementary black hole, but in fact it is the only information that *could* be encoded on an elementary surface.

I have already mentioned, in effect, that each elementary unit of surface which is considered to correspond with one binary unit of information in the semi-classical theory of black hole thermodynamics actually contains four of the units of surface which correspond with the smallest physically significant unit of distance (which is a Planck unit of distance). Why this should be the case has always remained unexplained. But in the context of the present semi-classical description of the degrees of freedom of matter particles which are under the influence of an elementary black hole, the fact that we need four units of area, or four discrete, elementary degrees of freedom to determine the state of each elementary particle present on the quantum gravitational scale no longer constitutes a mystery, because four microscopic parameters must be determined for each particle (one for the handedness, one for the sign of energy or action, one for the direction of propagation in time, and one for the sign of unified bidirectional charge) even though only the handedness of each particle contributes to determine the thermodynamic properties associated with the surface gravitational field of a macroscopic stable state black hole, as required by the semi-classical theory.

This conclusion, therefore, is not contradicted by the standard derivation

of the measure of black hole entropy, because under such conditions three out of each four units of missing information associated with what used to be an elementary unit of surface (which actually contains four elementary units of area) are irrelevant to the definition of the thermodynamic properties of the gravitational field and can actually be ignored. Indeed, the sign of energy of all particles that crossed the event horizon of a stable state black hole is fixed by the sign of energy of the object, while their direction of propagation in time only influences the microscopic properties of the field of interaction associated with the unified non-gravitational charge and the sign of bidirectional charge merely determines which component of this field encodes the information about the sign of charge, thereby leaving only the handedness of particles to be determined by the microscopic degrees of freedom associated with the surface gravitational field of a black hole. It is quite remarkable that such an exact quantitative result can be entirely derived from logical arguments made in the context of a semi-classical approximation. I believe that this conclusion, more than any other, illustrates the effectiveness of an unconventional approach such as the one I came to adopt for solving certain kinds of problems of particular importance in fundamental theoretical physics.

3.4 Large scale structure

I remember as a teenager, before I even learned about the existence of dark matter, having been deeply amazed and puzzled after reading in the newspaper that astrophysicists had determined that most of the visible matter in the universe, including our own galaxy, was located on the surface of giant voids of truly enormous proportions forming a bubble-like pattern in the matter distribution. I cannot say that I already expected back then that I would eventually be involved in developing a model that would help explain this troubling observation, but I did feel very strongly that this was something I needed to better understand. Anyhow, this stunning discovery and the mystery that initially surrounded it helped shape my early approach to the problem of gravitation in a way that turned out to be highly fruitful. What is truly remarkable is that the problem of voids has endured to this day as we kept discovering empty regions of increasingly larger sizes that still defy traditional explanations despite all the progress which was achieved in developing cosmological models that can more accurately reproduce those

features.

I believe that the introduction of negative energy matter will have an enormous impact on theories of structure formation. Indeed, what emanates from the results discussed in the preceding section is that the formation of structures in the visible matter distribution is accelerated by the presence of negative energy matter inhomogeneities, while the structures so produced in turn catalyze the formation of even larger inhomogeneities in the negative energy matter distribution. What needs to be emphasized in this regard is that given that in the early universe the average matter density was much larger than it currently is, then it follows that the underdensities present in the negative energy matter distribution had a much greater influence on positive energy matter, while the presence of voids in the positive energy matter distribution also had a significant influence on the formation of overdensities in the negative energy matter distribution, which through gravitational repulsion enhanced the formation of voids in the distribution of visible matter. Thus, negative energy matter is the source of additional gravitational instability which does not arise only from stronger gravitational attraction, but also from the gravitational repulsion exerted on visible matter by negative energy matter galaxies and clusters that conjointly develop as a result of the presence of primordial density fluctuations. Under such conditions the inhomogeneities which are present in the positive and negative energy matter distributions reinforce one another and accelerate the rate of structure formation.

But given that there are no direct interactions between positive and negative energy matter, one must conclude that even in the presence of primordial fluctuations in the density of negative energy matter, the spectrum of temperature fluctuations in the cosmic microwave background would not be significantly affected[7]. Thus, starting from the same relatively smooth initial matter distribution that is revealed by the low amplitude of CMB temperature fluctuations, we can expect higher density structures to develop at an earlier time, because negative energy matter is the source of additional gravitational instability. This allows to more easily reconcile the high level of development of present day inhomogeneities with the near perfect uniformity

---

[7]This does not mean, however, that there would be no effect at all on the CMB from the presence of variations in the density of negative energy matter at the epoch of last scattering. In fact, it is to be expected that the fluctuations of gravitational potential associated with those density variations would modify the spectrum of temperature fluctuations in a way that may help confirm the existence of negative energy matter.

of the temperature of CMB radiation. Those remarks would also apply to negative energy matter inhomogeneities themselves, which must be assumed to develop from an initial state as uniform as that of positive energy matter given that the specific densities of positive and negative energy matter are very similar initially and are submitted to the same constraints regarding the magnitude of primordial density fluctuations, as I will explain in the following sections of the present chapter.

The faster rate of development of large scale structures which can be expected to occur in the presence of negative energy matter would certainly help explain the most recent observations of huge voids in the galaxy distribution which computer simulations of structure formation based on the traditional cold dark matter model fail to reproduce. The case of voids is particularly interesting given that what triggers their formation is not merely gravitational attraction, but under the right conditions also the gravitational repulsion of negative energy matter overdensities which form inside the voids through gravitational attraction and which naturally produce more repulsive force on the surrounding positive energy matter, in addition to that which is provided by the voids themselves, thereby allowing them to grow even faster. Given that the average densities of positive and negative energy matter on the cosmic scale must be assumed to be very similar, while the average *specific* densities of positive and negative energy matter began to differ significantly only in the recent past (as reflected in the small positive value of the cosmological constant), it follows that we are allowed to expect that the invisible overdensities of negative energy matter concentrated in the voids have reached the same level of development as those observed in the positive energy matter distribution. These overdense structures would thus repel matter of positive energy sign and rapidly give rise to a cellular structure where sheets and filaments of positive energy galaxies surround large voids populated by similar, but invisible agglomerations of negative energy galaxies.

Now, as I previously explained, the gravitational field attributable to a large negative energy matter overdensity would be similar to that which is attributable to the void in the positive energy matter distribution in which it might be located. In such a context it follows that the presence of the overdensity would only enhance the gravitational repulsion of the void. I think that this is what explains that those gravitational fields were never identified as originating from the presence of gravitationally repulsive material. However, the additional contribution to the gravitational field of a void

that would be provided by the negative energy matter overdensity it may contain does have some distinctive effects, as it implies that smaller voids can exert an unexpectedly large gravitational repulsion. This is certainly a positive development given that it has been known for some time that certain voids in the positive energy matter distribution do exert larger than expected gravitational repulsion on galaxies located in their periphery, a phenomenon which had remained unexplained until now.

What must be retained is that the additional influence which is continuously being exerted on positive energy matter by both overdensities and underdensities in the negative energy matter distribution is significant enough to have given rise to structures which are already much more developed than those which are predicted by the conventional cold dark matter model. This is not a problem, but rather an advantage of the proposed approach, because it is no secret that the most recent observations have revealed the existence of structures whose existence at the present epoch has become increasingly more difficult to reconcile with conventional models of structure formation. It is obvious to me that such observations and the bubble-like pattern of the matter distribution in general can be much more easily explained if we allow for the existence of a parallel distribution of invisible, gravitationally repulsive, negative energy matter submitted to mutual gravitational attraction among particles of the same kind.

When gravitationally repulsive matter is present inside the voids in the visible matter distribution it is also easier to reconcile our theory of structure formation with those observations which show that there is a much smaller number of galaxies in the Local Void than is predicted by computer simulations, because any galaxy that would form in the void would rapidly be expelled to the periphery by larger than expected repulsive forces. Also, given that the density of negative energy matter in the Local Void would not be as low as it would in our galactic neighborhood, it follows that the missing mass effects attributable to negative energy matter underdensities would be more localized around those galaxies located nearer the void and this would have accelerated the formation of positive energy galaxies in this area. This may explain why a larger than expected number of very large galaxies in the Local Sheet are located on the periphery of the Local Void instead of in the more crowded areas where most of the visible matter is concentrated. In fact, this touches on a more general issue which is that the first large elliptical galaxies appear to have formed too early after the Big Bang for their creation to be easily explainable using conventional models. But

if we recognize that the presence of negative energy matter underdensities must have played a more important role on such a scale in the remote past, when the average density of negative energy matter was much larger and its distribution much more homogeneous, then this mystery can be explained quite straightforwardly.

The prediction that additional invisible structures must be present on a large scale in the negative energy matter distribution would also facilitate our understanding of certain observations which appear to show that coherent motions, involving a large number of galaxy clusters, are taking place above the Hubble motion. The magnitude of those large scale flows was unexpected from the viewpoint of the traditional cold dark matter model, but here again is a phenomenon which we should actually expect to occur when structures are present in the invisible distribution of negative energy matter that give rise to stronger gravitational fields. I believe that if those sources could be seen they would indeed allow to explain what causes the streaming motions. The fact that even the most recent detailed maps of the galaxy distribution fail to completely account for the matter whose gravitational field would be responsible for the motion of our own galaxy relative to the CMB clearly implies that additional sources of gravitational acceleration are involved in giving rise to this phenomenon. From my viewpoint those sources would simply be negative energy matter inhomogeneities.

In certain cases, however, the coherent motion is taking place on a scale so large that it was suggested that the inferred accelerations could never be attributed to the gravitational fields of structures located within the observable universe [32, 33]. If the existence of such bulk flows is confirmed it would mean that we are dealing with the same kind of anomaly as that which was revealed by certain observations which appear to show the existence of a privileged direction in the alignment of CMB temperature fluctuations or in the direction of rotation of spiral galaxies [34, 35]. It is likely that such features, if they are real, could only be explained as being the consequence of the existence of a very large scale polarization of the primordial distributions of positive and negative matter energy[8]. In the context of the explanation that will be provided in section 3.9 for the high level of homogeneity of the

---

[8]Such a structure is not forbidden by the empirical constraints which allow to rule out the existence of a very large spherical underdensity in the distribution of positive matter energy that could have been the source of the acceleration of universal expansion, given that the variation of density involved is much smaller and does not have the same unexplainable level of symmetry.

large scale distribution of matter energy, such a very large scale inhomogeneity would not be forbidden and if it does exist we can expect that it would be enhanced as a consequence of gravitational repulsion. Indeed, if the density of positive energy matter is slightly smaller on the average than the density of negative energy matter in a certain portion of the observable universe and slightly larger in the remaining portion, then we should expect to observe an alignment in the fluctuations of CMB temperature coming from opposite directions along a certain axis in space, given that such an inhomogeneity would give rise to a very large scale variation of gravitational potential along this axis which would contribute to further enhance the inhomogeneity of the primordial matter distribution by creating a force field that would accelerate positive and negative energy galaxies in opposite directions, thus giving rise to the observed very large scale bulk flows and perhaps also to the preferred direction of rotation of positive energy galaxies.

If this conjecture is valid, then in the context of my description of negative energy matter as being equivalent to missing positive vacuum energy an amazing conclusion would follow, which is that certain coupling constants and in particular the fine-structure constant α could vary along the axis in space relative to which the CBM temperature fluctuations are aligned. This would occur as a consequence of the fact that the coupling constants are affected by the virtual processes taking place in the vacuum, so that if energy is missing from the vacuum that would normally be carried by the virtual particles that interact with positive energy matter, then the renormalized value of the coupling constants could be reduced or increased in proportion to the amount of energy that is missing, which is proportional to the amount of negative energy matter that is present. If there are very large scale variations in the density of negative energy matter, it is possible that those variations of the coupling constants could become observable.

Concerning the electromagnetic coupling constant α in particular, it is known that the virtual processes responsible for vacuum polarization conspire to reduce the magnitude of the constant from its bare value and therefore if the effects attributable to those virtual processes are reduced, then α and the electric charge should be larger and this is what might occur in the presence of a background of negative energy matter. Thus, it may be appropriate to assume that α becomes smaller with time, given that the average density of negative energy matter decreases with time, which would imply that the 'constant' was larger in the far past when the amplitude of vacuum polarization effects was diminished as a result of the reduced level of activity

in the positive energy portion of the vacuum. This possibility appears to be compatible with the results of certain analyses of astronomical and terrestrial data (see [36] and [37] in particular) that show that the fine-structure constant is actually getting smaller with time and also varies across space along the direction of the observed cosmic bulk flows, just as we would expect based on the preceding arguments. Revealing the existence of variations in the value of α on smaller scale may remain a very difficult task, however, as fluctuations in the density of negative energy matter on intermediary scales were much smaller in the past, while the large overdensities in this matter distribution which must exist at the present epoch are probably confined to regions mostly devoid of visible matter.

Before concluding this section I would like to mention the existence of another remarkable astronomical phenomenon which might occur as a consequence of the presence of negative energy matter and which is also related to the more general issue of large scale structure given that its existence may actually have an impact on our assessment of the level of development of certain structures in the early universe. It involves an effect which might be called *repulsive gravitational lensing* and which is merely the counterpart to ordinary gravitational lensing that would be produced when the visible light from a distant source is gravitationally repelled while it travels through a negative energy matter overdensity on its way to our telescopes. In fact, such divergent gravitational lensing phenomena could also be caused by the presence of a positive energy matter underdensity located between a distant light source and the observer who measures its position in the sky, just like ordinary gravitational lensing can also be enhanced by the presence of underdensities in the negative energy matter distribution, superposed on the visible positive energy objects in the foreground, despite the fact that such a phenomenon is usually interpreted as being entirely attributable to the presence of positive energy dark matter.

What is interesting concerning those repulsive gravitational lensing phenomena in the present context is that they would distort the image of the background structures in such a way that the objects observed would appear to be more densely packed in space behind the invisible negative energy cluster located in the foreground. Thus, it could happen that background objects which were uniformly distributed at the moment when they released their light would appear to have already been assembled into clusters due to the divergent lensing to which their images are submitted. I believe that this could explain some observations which appear to show the presence of unex-

pectedly large quasar groups in the very distant past, which is problematic given that such clusters should not yet have had the time to develop at this early epoch. Some of those large quasar groups may very well be illusions arising from the presence of large overdensities of negative energy matter located in the light of sight of the instruments which measure the position of the very distant background objects. Just like ordinary gravitational lensing produces arcs of light, the repulsive gravitational lensing discussed here can be expected to produce blobs of light, which may in fact already have been observed in the X-ray spectrum. This would be a characteristic feature of negative energy matter cosmology which can therefore serve to confirm the validity of the basic hypotheses which enters the generalized gravitational model I have developed.

In face of the mounting difficulties we have encountered in recent years in trying to make sense of a growing amount of unexpected empirical results I think that the time has come to recognize that simply adjusting the free parameters of the cold dark matter model is no longer an adequate approach for addressing the challenges raised by the observed large scale features of our universe. But even if the words 'dark matter' are contained in the name of currently favored cosmological models it does not mean that rejecting those models requires completely abandoning the idea that invisible forms of positive energy may play a role in the development of large scale structures, because it remains that a certain phenomenon attributable to local variations of vacuum energy density can be expected to have consequences similar to those which were once attributed to conventional cold dark matter. Thus, I believe that what is required to make the current models more acceptable is merely an additional ingredient that would strengthen the gravitational forces responsible for sculpting the matter distribution on a large scale in ways which allows to appropriately describe certain phenomena which would otherwise remain unexplainable.

It is merely the fact that a void in the positive energy matter distribution is expected to exert a gravitational repulsion on the surrounding positive energy matter on a cosmological scale that prevents us from drawing the obvious conclusion that unseen matter must be present in the largest voids that may exert an even larger gravitational repulsion which greatly accelerates the rate of formation of those empty spherical structures and in the process allow large portions of the galaxy distribution to collectively reach relatively high velocities with respect to the cosmic microwave background. The early proposals that the largest voids might have formed as a consequence of explosive

processes that would have taken place in the early universe were thus based on the right intuition, but they failed because they did not involve gravitation as the repulsive force. It would therefore be the traditional reluctance to consider the possibility that gravitationally repulsive matter may exist, as well as the ignorance of the fact that such matter must necessarily be dark and be gravitationally attracted to itself, that would explain the difficulties we experience in trying to make sense of the most recent data regarding the processes that take place in our universe on a very large scale.

3.5   The flatness problem and matter creation

In the introductory section of this chapter I mentioned that there are two broad aspects to what I call the inflation problem which are the flatness problem and the horizon problem. Here I would like to discuss the first category of difficulty which will be shown to be indissociable from what is known as the problem of matter creation. Despite the commonly held belief that those problems have been solved by inflation theory I think that it is still important to understand the difficulties they raise for cosmology, given that the validity of inflation has not yet been definitely confirmed and even if there occurred an initial phase of accelerated expansion it may not necessarily produce the desired outcome. As I previously mentioned, the flatness problem arises from the fact that the present density of matter appears to be fixed to its critical value while we have no idea what the constraint is that would require such an extremely precise adjustment of parameters as would have to occur in the early stages of the Big Bang in order to produce the observed outcome. The problem is that if the faintest of deviation away from a critical rate of expansion had taken place at such an epoch, it would have given rise to a much larger deviation away from flatness at later times, while what we observe is a universe with an energy density that is still critical to a very good degree of precision. The truth, therefore, is that according to current knowledge, the Big Bang model, while mathematically consistent, is nevertheless incomplete, given that the initial conditions, it would seem, cannot be determined by the theory.

Of course this does not mean that we can't uniquely determine the rate of expansion at any time in the past by evolving the current state backward in time, which would actually allow us to predict that the density of matter has remained critical at all times in the past. Only, we cannot explain why

the current density ρ itself is fixed to its critical value ρ_c (associated with a density parameter $\Omega = \rho/\rho_c$ equal to 1) to such a high degree of precision. Thus, while relativity theory enables a positive energy observer to predict what the rate of expansion of the universe was at different times in the past given the current density of positive energy matter, according to the traditional approach this is only true in as much as the rate of expansion at the present time is empirically determined through a measurement of the Hubble constant H_0, but the model remains well defined for any value of Ω and H_0. Yet, I believe that there is much less freedom than is usually assumed in fixing the initial variation of the specific rates of expansion that give rise to the present specific densities of positive and negative energy matter. What I will now explain is that despite the conventional assumption to the effect that this initial condition is left unconstrained in the standard Big Bang model (without inflation) there does exist an unavoidable requirement for the current energy density of matter and vacuum to be very precisely equal to the critical value associated with a flat space from the viewpoint of both positive and negative energy observers.

One thing must be clear before we attempt to explain the current flatness of space on the cosmological scale and this is that there is an upper limit to the positive and negative contributions to the density of matter and vacuum energy. This means that space cannot continue to contract (in the past direction of time) beyond the point at which a maximum amount of energy of positive or negative energy sign is contained in every elementary unit of space. It would be incorrect to assume that the initial value of Ω cannot be determined due to the 'fact' that the initial density of matter is infinite in the very first instant of the Big Bang. Indeed, from a quantum gravitational viewpoint, there is no time zero at which the density of matter is infinite, only a minimum significant time at which positive and negative energy densities have a maximum, but finite magnitude. Given that in the context of my interpretation of matter as being equivalent to missing vacuum energy a maximum value of energy density is determined by the natural vacuum-stress-energy tensors associated with the upper limits of the positive and negative contributions to vacuum energy density, then this must be assumed to be the maximum magnitude of the positive and negative contributions to the density of matter and vacuum energy in the state that emerges from the initial singularity.

What needs to be explained, therefore, is merely why it is that the rate of expansion of space did not begin to differ from its critical value immediately

after the universe emerged from this state of maximum positive and negative energy densities that is uniquely determined by the natural scale of quantum gravitational phenomena. The initial positive and negative densities of non-gravitational energy are not arbitrary, but the problem is that there is too much freedom in fixing the early variation of the rate of expansion which determines the average density of matter at all later times. From the conventional viewpoint it would appear that the early variation of the rate of expansion that gives rise to a flat space at the present time is merely one alternative among an enormous range of possibilities. What I will explain, however, is that while the current value of gravitational potential energy for the universe as a whole (which is fixed by the present average densities of positive matter and vacuum energy) and the currently observed kinetic energy of expansion (which is determined by H_0) appear to constitute free parameters of the standard model of cosmology, they are not really independent variables in the context where energy must be null for the universe as a whole. In fact, under such conditions, the early variation of the rate of expansion measured by a positive energy observer must be adjusted not just to a level of precision that would allow space to keep expanding until the present epoch, but to such an extent that space can be expected to remain perfectly flat on the largest scale for an arbitrarily long time. I will show that this constraint can only be fulfilled when negative energy matter is assumed to be present in the universe alongside ordinary positive energy matter.

In the context of the model I have proposed to integrate negative energy matter to gravitation theory it may seem like the presence of such negative energy matter would change nothing to the conclusion that flat space is an unlikely possibility for the present state of the universe, because a uniform distribution of negative energy matter exerts no influence on the gravitational dynamics of positive energy matter on the largest scale, for reasons I have explained in section 1.6. The present specific density of negative energy matter would in fact be independently subjected to the same excess of freedom as affects that of positive energy matter, given that the variation of the specific rate of expansion of negative energy matter is determined only by the density of matter with the same energy sign and it would appear that this expansion rate could vary as freely as the specific rate of expansion of positive energy matter initially. In any case, if space was negatively curved from the viewpoint of a negative energy observer, this would not merely be a consequence of the fact that the energy of matter that determines the expansion rate measured by such an observer is indeed negative, as if negative

energy matter could accelerate its own specific rate of expansion through gravitational repulsion, in the way we would expect from a traditional viewpoint, because as I explained in section 1.4 negative energy or negative mass matter does not exert a gravitational repulsion on matter with the same energy sign. Thus, in principle, space could just as well be positively curved and closed from the viewpoint of a negative energy observer, because the property of gravitational attraction or repulsion is not an absolute feature of matter with a given energy or mass sign. Yet, despite this state of affairs, it turns out that the presence of negative energy matter is in fact required (as I mentioned above) to explain why it is that we are allowed to expect that space should be perfectly flat from the viewpoint of a positive energy observer.

Although the alternative solution I will propose to the flatness problem is quite simple, it was actually one of the results which I had the most difficulty deriving among those that figure in this report. Part of the difficulty arose from the fact that there are conflicting accounts of what constitute the many contributions to the energy budget of the universe and how their magnitudes may vary as a function of the values assumed by various physical parameters. Thus, while I always had the intuition that in the context where the presence of negative energy matter cannot be ignored, a natural solution to the flatness problem might become possible once we recognize the necessity to appropriately apply the principle of energy conservation to the Big Bang, it was not clear which contributions could balance one another out exactly in order to produce a universe out of nothing. But when I finally figured out what the various contributions to the energy budget of the universe are in the presence of negative energy matter, and which must be considered independent from which others, and which would need to have the same magnitudes in the initial Big Bang state, then it became clear that under such conditions space must actually expand at precisely the critical rate when we require the energy to be null for the universe as a whole. Before I explain why it is exactly that applying a constraint of energy conservation to the Big Bang may have such far-reaching implications, however, I would like to describe what the motives are that justify assuming that the energy of the universe must in effect be null.

I already discussed the importance and the unavoidable character of the constraint imposed by the requirement of relational definition of physical attributes in the preceding two chapters. Basically, what must be understood concerning the problem at hand is that the total energy of the universe

constitutes one such property which definitely cannot violate the rule that it be characterized in a purely relational way. What is implied by such a requirement is that even if the Big Bang was not considered to constitute a creation event at which any conserved physical quantity must be created out of nothing, from the viewpoint of an observer of any energy sign the universe would still need to have a vanishing total average energy density. Indeed, one might argue that the requirement of invariance in time of conserved physical quantities does not apply to such a singular event as the Big Bang at which time itself may come into existence, or alternatively that the Big Bang does not even constitute an absolute beginning to time given that evolution could perhaps be continued to times before the initial singularity if a 'Big Bounce' occurs, as was proposed by certain authors in a quantum gravitational context. But when we recognize the unavoidable nature of the constraint of relational definition of the physical attribute of energy it emerges that the universe as a whole cannot have a non-zero energy, even if the Big Bang does not constitute a creation event at which any conserved quantity must be created in equal positive and negative amounts.

This conclusion simply follows from the fact that if it was possible to measure a non-zero value for the energy of the universe as a whole, then this value would have to be either positive or negative and this would allow the particular direction of time relative to which this positive or negative energy would propagate to be singled out as an absolutely defined direction, in the exact same way a non-zero momentum for the universe as a whole, arising from a collective motion of positive energy matter relative to negative energy matter, would allow to single out a particular direction in space as being that along which this positive or negative momentum is directed. Here the fact that there exist both positive and negative energy particles propagating forward in time is no different from the fact that there may exist particles with both positive and negative momenta propagating in one and the same direction of space. If the positive and negative energy matter distributions had a non-zero total momentum as a result of being in motion relative to one another on the average, then no reference system would exist relative to which this momentum would vanish, unlike would the case if only positive energy matter was assumed to be present. What I'm suggesting is that the same is true for the non-gravitational energy of the universe, which must therefore remain null under all circumstances (as becomes possible in the presence of negative energy matter) if one is to avoid giving preferred status to one particular direction of time for the universe as a whole. Thus, I

believe that if a non-zero value for the total energy of matter was allowed to develop it would necessarily give rise to a compensating contribution by vacuum energy that would leave null the total measure of non-gravitational energy experienced by positive and negative energy observers.

Indeed, based on the developments introduced in section 3.2 it would appear that when the average, specific density of negative energy matter is growing relative to that of positive energy matter as a consequence of the emergence of a difference between their specific rates of expansion (the rates of expansion experienced by negative and positive energy observers respectively), the ratio of the average densities of positive and negative energy matter determined by a positive energy observer must remain invariant, because the density of negative energy matter measured by such an observer is modified by the same metric conversion factor which fixes the density of vacuum energy, while the density of vacuum energy must grow in proportion to the magnitude of the divergence between the scale factors experienced by opposite energy observers. As a result, any variation of the average, specific density of negative energy matter relative to that of positive energy matter remains unobservable for a positive energy observer, which means that if the total density of matter energy was null initially, then it would remain so as expansion takes place and the same conclusion would apply to the average matter densities determined by a negative energy observer.

The situation with momentum is a little different, as any variation in the momentum state of positive energy matter relative to that of negative energy matter can only develop locally, and in such a case the usual conservation laws which apply when positive and negative energy matter interact provide sufficiently strong a constraint to alone prevent a non-zero momentum to develop in one or another direction of space. For those reasons, I believe that the commonly held opinion to the effect that it may not be absolutely necessary to require the universe to have a null value of momentum or a null value of energy, as would appear necessary when the principle of conservation of energy applies to the creation process which occurred in the first instants of the Big Bang, cannot be justified. The fact that by taking a different stance I will achieve significant progress in describing the early stages of the universe's expansion will serve, I hope, to vindicate the legitimacy of my viewpoint.

Now, what most people already recognize concerning the energy content of the universe is that for a flat universe with a zero cosmological constant the negative gravitational potential energy of positive energy matter and ra-

diation is balanced by the positive kinetic energy of expansion of this matter. When that is not the case then an additional amount of gravitational potential energy is present that is attributable to the gravitational field itself (or the curvature of space) and that tends to become dominant very rapidly (regardless of whether it is positive or negative) as space expands, because while the gravitational potential energy of matter decreases in inverse proportion to the volume, the gravitational potential energy associated with the curvature of space decreases as the inverse of the surface enclosing that volume. What may be difficult to understand is the fact that the kinetic energy of expansion is actually a property of the expanding space, which means that it must be considered an energy of the gravitational field itself and not really an energy of matter, despite the fact that the sign of this energy varies as a function of the sign of energy of the observer which is assessing its value. Indeed, the initial value equation for a homogeneous and isotropic universe, which is derived from the general relativistic gravitational field equations under the condition that energy is conserved for the universe as a whole is usually written as

$$E = K + V(a) = \left(\frac{1}{a}\frac{da}{dt}\right)^2 + \left(\frac{-8\pi\rho}{3} - \frac{\Lambda}{3} + \frac{k}{a^2}\right) = 0 \qquad (3.3)$$

where E is the gravitational (potential) energy of the universe, K is the kinetic energy of expansion and $V(a)$ is the Friedmann potential as a function of the scale factor $a(t)$ in the presence of a cosmological constant Λ for a universe with an average matter density ρ. Here the spatial curvature parameter, which I redefine as $-k/a^2$ and which is always precisely equal to zero for a flat universe, appears as just one particular (reversed) contribution to the Friedmann potential, but when it is possible to assume that the magnitude of the cosmological constant was negligible initially this equation can be rewritten as

$$E_k = K + U(a) = \left(\frac{1}{a}\frac{da}{dt}\right)^2 - \left(\frac{8\pi\rho}{3}\right) = \frac{-k}{a^2} \qquad (3.4)$$

which clearly shows that the spatial curvature parameter is the outcome of the imperfect cancellation of the gravitational potential energy of matter by the kinetic energy of expansion.

Thus, whenever the gravitational potential energy of matter $U(a)$ is not matched by a kinetic energy of expansion K that's exactly its opposite, the

gravitational potential energy E_k attributable to the gravitational field itself, which is given by $-k/a^2$, is not zero and contributes to alter the expansion rate. If k is positive this excess of gravitational potential energy is negative, which means that the source of the gravitational field must then have positive energy, as must be the case when the negative gravitational potential energy of matter itself contributes predominantly to determine the gravitational field, while when k is negative there is a positive excess of gravitational potential energy, which means that the source of the gravitational field has negative energy, as is the case when the positive kinetic energy of expansion (which is also an energy of the gravitational field) contributes predominantly to determine the gravitational field of the universe. The gravitational potential energy E_k associated with the present value of the curvature parameter $-k/a^2$ must therefore be considered to consist of a residual measure of energy which could in principle assume any positive, negative, or null value depending on the current value of the scale factor and on whether k is equal to -1, $+1$, or 0. There is no a priori reason, however, to assume that the measure of gravitational potential energy associated with the curvature of space on the cosmological scale should be the same for positive and negative energy observers at the same epoch, because the kinetic energy of expansion varies as a function of the rate of expansion, which is an observer dependent quantity in the context where, as I explained in section 1.6, only the average density of positive energy matter contributes to determine the gravitational field that influences the expansion rate measured by a positive energy observer, while in principle a negative energy observer could measure different values for the average density of matter and the rate of expansion, for reasons I discussed in section 3.2.

It must be clear that even though it is usually assumed that the initial value equation expresses the requirement of gravitational potential energy conservation for the universe as a whole in a general relativistic context, what the original form of the equation really means is that when an additional term, which is provided by the negative of the spatial curvature parameter $-k/a^2$, is added to the equation that would otherwise express the nullity of gravitational potential energy, then the gravitational energy of the universe can be conserved even in those cases where it would not really be null initially, but it does not really amount to require that the universe comes into existence with zero gravitational energy. What equation (3.4) means is that once it is assumed that the cosmological constant Λ is negligible initially, then it is only when the free parameter $-k/a^2$ associated with the

curvature of space is zero that the positive kinetic energy of expansion K can balance the negative gravitational potential energy $U(a)$ attributable to the presence of positive energy matter. The true measure of gravitational energy for the universe as whole, therefore, is really that which is associated with the curvature of space (which would justify that we refer to this energy as the actual gravitational energy of the universe) and it is only when this energy is null that the gravitational field does not contribute energy on the cosmological scale. But it is usually assumed that this curvature parameter can also be positive or negative and the universe be positively or negatively curved, so that the degree of curvature at any given time would depend on the initial value of the kinetic energy of expansion when the density of matter and radiation was maximum. It must be acknowledged, however, that from the viewpoint of positive energy observers at least, space does have a flat geometry to a relatively good degree of precision and this means that there must be a reason why the curvature parameter has a null value.

I believe that what allows the value of gravitational potential energy E_k associated with the spatial curvature parameter to be null for an expanding zero energy universe is the fact that the gravitational potential energy of matter experienced by a positive energy observer can be arbitrarily large even when negative energy matter is present and the total energy of matter itself is null. Indeed, when negative energy matter is present a flat universe can actually have zero energy, despite the fact that from a traditional viewpoint it would appear that if the energy contained in the gravitational field we experience was null (as would occur if the negative gravitational potential energy of matter was compensated by the kinetic energy of expansion) the energy of the universe would still be positive (because the energy of matter would not cancel out). It is only from a traditional perspective that it would appear impossible to require our flat universe to have zero energy. In the absence of negative energy matter the universe would actually need to have a positive curvature in order to have zero energy, because only then could the negative energy contained in the gravitational field compensate the positive energy of matter (while the gravitational field of a negatively curved universe would contribute more positive energy, as the positive kinetic energy of expansion would overcompensate the negative gravitational potential energy of positive energy matter to provide a positive gravitational potential energy E_k). In fact, it seems that it is only for a closed universe that does not expand at all that the positive energy of matter could be entirely compensated by the residual energy of the gravitational field in the initial state of maximum

matter density, because according to certain accounts, when the density is that high, the gravitational potential energy is actually equal in magnitude to the energy of matter. Again, however, the problem is that the universe is not highly curved, but in all likeliness almost perfectly flat.

At this point it is important to mention that the idea that the energy of the universe should perhaps be required to be null is not new. Thus, it was once suggested [38] that the universe could fluctuate into existence if the positive energy of matter could be compensated by its negative gravitational potential energy, at least in the very high density of a primordial state. The problem was that it appeared that such a highly curved universe could never be produced as a fluctuation out of nothing, because if it actually has zero energy it would only be allowed to expand for a very short period of time before immediately recollapsing back to the vacuum. Creation out of nothing was eventually salvaged from this severe failure by assuming that once in a while inflation may occur when a universe is fluctuating out of the vacuum, which would enable its expansion rate to start growing exponentially thereby giving rise to a flat space which would keep expanding indefinitely.

I will not immediately discuss any motives we may have to resist appealing to inflation in order to solve the problem of creation out of nothing or indeed any other problem, but given that very early on I chose to explain known facts with principles which are themselves known to be valid with absolute certainty (even if certain consequences of applying those fundamental principles may not yet be recognized as unavoidable), then I will propose a different solution to the problem of creation out of nothing. In order to proceed in this direction, however, one must first acknowledge that if the negative gravitational potential energy of matter exactly balances the positive kinetic energy of expansion for a flat universe, then this gravitational potential energy cannot also balance the positive energy of matter itself, as earlier proposals required assuming. This does not mean that the magnitude of the gravitational potential energy experienced by a positive energy observer cannot be equal to the magnitude of positive matter energy initially, only that this is not an appropriate and sufficient condition for obtaining a zero energy universe. In fact, as I mentioned above, it does appear, according to certain accounts, that in the initial Big Bang singularity (or indeed any other singularity) the positive energy of matter is equal in magnitude to its negative gravitational potential energy and this is precisely the reason why it was so difficult for me to realize that it is not appropriate to merely require the gravitational potential energy to compensate the energy of matter

in order to obtain a universe with zero energy.

What I have realized is that in a zero energy universe any residual measure of gravitational field energy associated with the initial value of the spatial curvature parameter $-k/a^2$ determined using the metric properties of space experienced by positive energy observers must necessarily balance the residual energy of matter obtained by adding the opposite contributions of positive and negative energy matter. Now, if the curvature parameter is null, like gravitational energy itself, in the case of a flat universe (for which the kinetic energy of expansion experienced by an observer with a given energy sign precisely balances the gravitational potential energy of the matter with the same sign of energy), then it can only mean that in such a case the energy of matter must itself add up to zero. Normally that would not be possible, because only an empty universe would have a null, average density of matter energy. But in the presence of negative energy matter a high density universe can actually have a null matter energy, as long as the average densities of positive and negative energy matter have exactly the same magnitude. The idea that the negative energy of matter could compensate its own positive energy may seem problematic in the context where I have explained (in section 1.11) that those two energies are conserved independently from one another under most circumstances. One must remember, however, that we are dealing with the Big Bang here.

In section 1.9 I have mentioned that under conditions of very short time duration it is to be expected that pairs of opposite action particles can be created out of nothing without violating the constraint of energy conservation. The density of matter that is continuously being created and annihilated in such a way is determined by the natural scale of quantum gravitational phenomena and therefore actually constitutes a maximum density. Normally the opposite action particles so created immediately annihilate back to nothing, because their energies are sufficiently large to allow them to interact with one another (even if only indirectly). But under conditions where the initial rate of expansion of the universe is sufficiently high (as we may assume to be the case for a flat universe for which the kinetic energy of expansion measured by a positive energy observer would balance the very large gravitational potential energy of the positive energy matter that is present on the quantum gravitational scale as a consequence of the existence of such opposite action pair creation processes), then it should be possible for matter to be permanently created, because the high expansion rate involved would prevent the created particles from annihilating back to the vacuum as they

normally would. In fact, I previously remarked that this may be the only way to explain the presence of matter in our universe in the context where processes of annihilation to nothing are also occurring on the quantum gravitational scale that would tend to eliminate any opposite action particles that would already be present in the vacuum. But if opposite action pair creation processes are actually responsible for the presence of matter in the initial Big Bang state, then it means that equal numbers of positive and negative energy matter particles are necessarily produced, which may allow the energy of matter to be null for the universe as a whole.

It must be clear, however, that even if we were to assume that there are as many positive action particles as there are negative action particles in the initial state of maximum matter density, in principle it would still be possible for the average density of positive matter energy to be larger or smaller than the average density of negative matter energy, even in a universe with zero energy. In the absence of an appropriate constraint this would, in effect, be allowed as long as the differences between the positive and the negative energies of matter are compensated from the viewpoint of a given observer by the energy of the gravitational field associated with the curvature of space, which is determined by the rate of expansion measured by that observer (because while negative matter energy can compensate positive matter energy, only the gravitational field experienced by a positive energy observer can contribute to cancel out any non-zero, average energy density of matter determined by such an observer). Under such conditions the magnitudes of the positive and the negative contributions to the energy of the universe could be equal initially, even if the average energy densities of positive and negative energy matter were not themselves equal and therefore the total energy could in principle be null regardless of the amount of energy contained in the gravitational field. It may therefore seem like a condition of null energy for the universe as a whole and the creation of all matter out of nothing do not provide sufficiently strong a constraint to necessarily give rise to a flat universe. But, in fact, I came to realize that this condition is much more constraining for gravitational energy and the rate of expansion than may at first appear to be the case and that it actually allows to predict that the geometry of our universe must be flat on the largest scale.

It is important to point out, first of all, that the nullity of the energy of matter cannot be fixed as an independent consistency requirement, because that would require assuming that there cannot even be local fluctuations away from this zero energy for matter, while this is required to explain the

observed inhomogeneities present in the initial distribution of matter energy on a scale larger than the cosmic horizon. But in the absence of such a constraint, local fluctuations above or below the average zero value of matter energy density could in effect be present in the initial Big Bang state, even if the average densities of positive and negative matter energy were required to cancel out so as to allow the zero energy universe to have a flat geometry, as long as there is in effect as much overdensity as there is underdensity in the positive and negative energy matter distributions on a sufficiently large scale. Such fluctuations in matter energy would simply need to be compensated by local variations in the kinetic energy of expansion above or below the value associated with a critical expansion rate. Therefore, fluctuations would be allowed, even in a maximum density state, given that the variations of gravitational field energy would actually compensate the variations in matter energy and maintain the positive and negative energy densities (of matter and gravitational field together) at their maximum value. If there was less positive than negative matter energy in a certain location initially, then there would simply need to be more positive gravitational energy and therefore more positive kinetic energy of expansion for positive energy matter and less negative kinetic energy of expansion for negative energy matter.

Thus, inhomogeneities could be present in the initial distribution of matter energy, even if the density of positive energy particles (the number of positive action particles in a volume of space) was required to everywhere equal that of negative energy particles in the context where those particles are assumed to be produced by pair, because *locally* at least the nullity of energy can arise from a compensation between the energy of matter and the energy of the gravitational field. Indeed, a local variation in the energy of the gravitational field (attributable to a local variation of the kinetic energy of expansion above or below the value associated with a critical expansion rate) can be made to compensate any local difference between the magnitude of the density of positive matter energy and that of negative matter energy, just like the global measure of gravitational field energy which is attributable to the difference between the observer dependent gravitational potential energy of matter and the observer dependent kinetic energy of expansion could in principle compensate any difference between the magnitude of the average cosmic densities of positive and negative matter energy. However, in section 3.9 I will explain that a certain unavoidable constraint actually limits the amplitude of those fluctuations in the initial state and therefore it cannot be expected that there would occur large deviations from zero gravitational

energy locally if this condition is also obeyed globally.

But even if local fluctuations in the density of matter energy are clearly unavoidable it remains to explain why it is that such a compensation of matter energy by gravitational energy is not allowed to take place on a global scale, as required if space is to be flat for the universe as a whole. Indeed, as I mentioned above, if a residual gravitational energy associated with the spatial curvature parameter $-k/a^2$ could also compensate a difference in the magnitude of the initial, average energy densities of positive and negative energy matter on a global scale, then it should be possible for the magnitude of the kinetic energies of expansion experienced by positive and negative energy observers to be larger or smaller than the magnitude of the gravitational potential energies of their associated matter. Under such conditions the rates of expansion would no longer need to be critical, even in a zero energy universe. It is certainly true that a homogeneous distribution of negative energy matter exerts no influence on the specific expansion rate of positive energy matter which determines the kinetic energy of expansion measured by a positive energy observer, but this is significant merely in the sense that only the energy of the gravitational field perceived by a positive energy observer can contribute to the energy budget that must add up to zero on a global scale from the viewpoint of such an observer. In the context where all matter is created out of nothing as opposite action pairs it is still necessary to assume that both the positive and the negative energy of matter contribute to the total measure of energy effected by a positive energy observer.

Indeed, negative energy matter also does experience the expansion rate measured by positive energy observers when its motion is described using the metric properties of space determined by such an observer and therefore its presence must also be taken into account in balancing the energy budget associated with a positive energy observer, even if it does not directly influence the rate of expansion that is measured by such an observer on the scale at which the matter is homogeneously distributed. We may, in fact, consider that the way by which negative energy matter does contribute to determine the gravitational field experienced by a positive energy observer on a global scale is precisely by reducing the energy of matter, which allows the negative portion of the energy of the gravitational field to itself be reduced when the total energy is required to be null. Anyhow, either the negative energy of matter remains totally uncompensated by the energy of the gravitational field associated with a positive energy observer, in which case this gravitational field energy would alone need to compensate the positive energy of

matter, which would imply that there can be no contribution by a positive kinetic energy of expansion, so that the universe should not expand at all, or the negative energy of matter must be totally compensated by the same gravitational field energy, along with that of positive energy matter, in which case expansion is allowed to occur, but we must explain why the total average density of matter energy was initially so close to zero that the energy of the gravitational field (associated with the global curvature of space) was itself required to be perfectly null. Clearly, the second option is the only one that *could* be viable and therefore I will concentrate on explaining why the average energy of matter which balances the energy of the gravitational field for the universe as a whole cannot be as arbitrarily large as one might otherwise expect.

One thing must be clear first of all and it is that if space was positively curved and closed from the viewpoint of a positive energy observer, it would mean that it is negatively curved and open from the viewpoint of a negative energy observer. Indeed, the energy of the gravitational field of a universe that would be positively curved from the viewpoint of a positive energy observer would be negative and could therefore only compensate an excess of positive matter energy (through a reduction of the positive kinetic energy of expansion). But while it is true that even from the viewpoint of a negative energy observer an excess of positive matter energy would require the contribution of a gravitational field with negative energy, such a gravitational field would be associated not with a smaller positive kinetic energy of expansion, but with a larger negative kinetic energy of expansion and a higher than critical expansion rate which would actually give rise to an open universe. If the total energy of matter was instead negative, as would occur if negative energy matter particles contributed more energy than positive energy matter particles on the average, then the opposite would be true and the universe would appear to be closed for a negative energy observer and open for a positive energy observer. Now, while those two mutually exclusive configurations may appear as merely consisting of two additional possibilities, no different from the case where the average energy density of matter happens to be null, just like the energy of the gravitational field, there is actually a very important distinction between the case of a flat universe and that of the curved space configurations. This essential difference has to do with the fact that in the first case the universe would be open from both the viewpoint of a positive energy observer and that of a negative energy observer, while in all the other possible cases it seems that the universe would need to be open

for an observer with a given energy sign and closed for an observer with the opposite energy sign.

I believe that if the average density of positive matter energy must very nearly compensate the average density of negative matter energy in the initial Big Bang state, even though local fluctuations away from the zero energy of matter are allowed to be present to a certain extent (as long as they are compensated by opposite local fluctuations in gravitational energy), it is precisely because, in the absence of any other contribution to the energy budget, if matter energy was not null, then space could not be flat and open from the viewpoint of all observers. If an excess of positive or negative gravitational energy was allowed to compensate an excess of negative or positive matter energy (respectively) on a global scale, then this excess gravitational energy would give rise to a universe which would be open for one observer and closed for an observer with opposite energy sign. But given that the difference between the volume of a closed universe and that of an open universe would in principle be infinite it follows that such a configuration would be characterized by an arbitrarily large positive or negative density of vacuum energy. Indeed, from the viewpoint of the developments discussed in section 3.2 it would follow that if gravitational energy was positive and the universe was open from the viewpoint of a positive energy observer and closed from the viewpoint of a negative energy observer, as would appear to be required if it is to compensate a negative total density of matter energy, the density of vacuum energy should be positive with a maximum amplitude, while if the opposite is true and gravitational energy is instead negative, as would appear to be required if it is to compensate a positive total density of matter energy, then the density of vacuum energy should be set to its maximum negative value right at the Big Bang.

The problem is that this just cannot be the case, because vacuum energy also contributes to the positive or negative density of energy which must be canceled out by the energy of the gravitational field. Thus, if the average density of positive matter energy was smaller than that of negative matter energy and space was open from the viewpoint of a positive energy observer, the positive density of vacuum energy would rather simply compensate the positive matter energy deficit, which would *necessarily* make space flat from the viewpoint of all observers. If it was instead the average density of negative matter energy which was smaller than that of positive matter energy and space was closed from the viewpoint of the same positive energy observer, then a negative density of vacuum energy would compensate the negative

energy matter deficit and again make space flat from the viewpoint of both positive and negative energy observers. Therefore, even if the density of positive matter energy was null and that of negative matter energy was maximum, the density of positive non-gravitational energy would not really be smaller than that of negative non-gravitational energy, but would also be maximum due to the contribution of positive vacuum energy. What is important to understand is that under such circumstances the total energy would still be null for the universe as a whole, because the maximum positive density of vacuum energy would cancel the maximum negative density of matter energy and this would always be observed regardless of the actual configuration. In practice it is not gravitational energy that compensates matter energy on the cosmic scale, but really vacuum energy (as if the scale factors where not equal despite flatness, which is not forbidden). But the crucial point is that such a compensation must necessarily take place whenever the total density of matter energy is not perfectly null, as I have just explained.

It is usually recognized, however, (as I mentioned in section 3.2) that if vacuum energy density is too large initially, then it would prevent the development of an observer. From my viewpoint, therefore, the weak anthropic principle would not allow a configuration where a large disparity exists between the initial density of positive matter energy and that of negative matter energy, even if a small non-zero value for the energy density of matter would not make the density of vacuum energy *maximum* to begin with (as I had assumed at some point before deriving the above argument). Only a universe with very precisely, but not necessarily perfectly balanced contributions to the energy of matter is allowed to be experienced as a long lasting process by a physical observer that is part of that universe when it is appropriately required that the universe itself has null energy. The fact that the energy contained in the gravitational field on a global scale must also be null, independently, would then appear to be an additional consistency requirement that is naturally fulfilled in the above described context as a result of the variable contribution to non-gravitational energy that is provided by the vacuum, which always conspires to produce a flat geometry from the viewpoint of all observers, regardless of any changes to the total density of matter energy. The rate of expansion must always be critical, but it is only when the critical density of energy is not contributed for the most part by the vacuum that an observer is allowed to be present at some point in the universe to measure any value for the total density of matter energy.

Anyhow, one must conclude that the kinetic energy of expansion de-

termined by a positive energy observer would always precisely compensate the gravitational potential energy attributable to positive matter energy and vacuum energy, while the kinetic energy of expansion determined by a negative energy observer would always compensate the gravitational potential energy attributable to negative matter energy and vacuum energy (whatever its energy sign). When this is properly understood it becomes clear that the 'extra' principle which is required in order to fix the expansion rate of our universe to its critical value is nothing else but the requirement of relational definition of physical attributes applied to the energy of the universe. In the context of the generalized gravitation theory introduced in the first chapter of this report and given the interpretation that was proposed in section 3.2 for the cosmological term, this constraint actually allows to determine which solution of the gravitational field equations is the appropriate one for a description of the expanding universe. It is, therefore, by applying this very basic principle, in the context where it is recognized that negative energy matter must also contribute to the universe's energy budget, that it becomes possible to explain not only why there is expansion, but why it is that the rate of this expansion is still critical, even long after the Big Bang. Space is flat and the rate of expansion remains critical, because that is the only possible configuration that is allowed in the context where any difference between the scale factors determined by observers of opposite energy signs contributes to determine the average value of vacuum energy density that is experienced by those two types of observers, which in turn contributes to determine the curvature of space.

Now, given that the amplitude of temperature fluctuations in the cosmic microwave background provides strong observational constraints on the magnitude of initial inhomogeneities in the distribution of negative matter energy, which cannot therefore be much larger than that of the inhomogeneities present at the same time in the distribution of positive matter energy and given that the magnitude of the inhomogeneities present at the epoch or recombination is dependent on the density of matter energy, then it is possible to conclude that measurements do confirm that the magnitudes of the average densities of positive and negative matter energy were very similar initially, which allows the cosmological constant to itself be nearly perfectly null at the Big Bang and consequently also at the present epoch, as necessary for the existence of a physical observer.

What must be retained from all this is that if it was not for the fact that the presence of a homogeneous distribution of negative energy matter exerts

no influence on the expansion rate of positive energy matter (as explicitly stated in the formulation of principle 6 from section 1.14 and for reasons I explained in section 1.6), then, even if the energy of matter was null, it would not be possible to conclude that the initial expansion rate must be the critical rate associated with the density of positive matter energy, because under such conditions the gravitational potential energy of matter and vacuum that would need to be balanced by the kinetic energy of expansion would actually be zero (because the total density of matter and vacuum energy that would determine the strength of the gravitational field would itself be null), which means that the kinetic energy of expansion would also be zero and the universe should not expand at all. But if the requirement of energy conservation did not apply to the gravitational field and the universe did expand, as we would normally assume, then the expansion rate would not be submitted to any deceleration and the universe would explode like a negatively curved universe with a null matter density. The independence of the expansion rates of positive and negative energy matter from the presence of matter with an opposite energy sign, which follows from my description of negative energy matter as consisting of voids in the positive energy portion of the vacuum, is therefore an essential ingredient of the alternative solution to the problem of flatness that is proposed here. This condition is especially constraining in the context where the distribution of matter needs to be assumed highly homogeneous on the largest scale (for reasons I will explain in section 3.9) so that there cannot even exist significant local perturbations of the rate of expansion of matter with a given energy sign by matter with an opposite energy sign on such a scale.

It is only after I realized that the presence of negative energy matter does not contribute to determine the gravitational potential energy attributable to the presence of matter and vacuum energy, which is compensated by the kinetic energy of expansion measured by positive energy observers in the case of a universe with an overall flat geometry, that I was able to understand that despite what is usually assumed it is in effect not only the current *variation* of the specific rate of expansion of positive energy matter which is determined in part by its energy density, but actually also the current specific rate of expansion itself. It took me a certain time to recognize that the variation of the rate of expansion must indeed be considered to depend on the density of matter energy, as most people may consider obvious, but my questioning has allowed me to realize that the relation which exists between the rate of expansion and the density of matter energy is actually much more

constraining than is usually believed.

As a result, I'm allowed to conclude that even in the absence of inflation it is not necessary to assume that the present density of positive matter and vacuum energy is critical purely for aesthetic reasons, because in fact it is possible to explain why the universe is so beautifully balanced when one recognizes the necessity to properly apply the requirement of relational definition of physical attributes to the energy of the universe as a whole, which requires energy to be conserved even under the extreme conditions characteristic of the initial Big Bang state. Furthermore, it appears that it is the fact that an observer can only measure a value of vacuum energy density that is compatible with the conditions of her own existence that explains that it is not merely the total energy content of the universe that is observed to be very precisely null, but to a good degree of precision, also, the total energy of matter. The flatness of space is not merely a possibility and the zero energy of matter a coincidence, as would be the case if it was inflation theory that explained flatness. Instead, both conditions constitute basic consistency requirements that must be satisfied by any viable cosmological model.

It is important to mention that the processes of matter creation discussed above do not occur in an extended vacuum that emerges from a prior state of accelerated expansion, as would be the case with traditional creation out of 'nothing' scenarios developed in the context of eternal inflation. Here matter creation is rather allowed to constitute a true beginning for the history of the universe from the viewpoint of unidirectional time (even if bidirectional time itself may extend past the initial singularity and in such a way give rise to a time-symmetric configuration for which creation still occurs only once). The fact that in the context of the approach discussed above this history can actually begin with a state of maximum matter density, even when the average density of matter energy is required to be null, will later be shown to be a highly desirable feature given that it allows the elaboration of a more consistent solution to the horizon problem.

When I will discuss this important issue in section 3.9 I will explain what justifies assuming that the energy of the matter that is produced by the processes of opposite action pair creation which naturally occur on the time scale characteristic of quantum gravitational phenomena (the Planck time) is so homogeneously distributed that despite its very high density (also characteristic of the quantum gravitational scale) it does not form macroscopic event horizons. In the context where the initial distribution of matter energy

is uniform to a very high degree and the *local* rates of expansion of positive and negative energy matter only vary in such a way as to allow the kinetic energy of expansion to compensate any difference between the amplitudes of their opposite energy densities, as I'm here assuming, then the matter distribution remains homogeneous and the expansion rate isotropic on the largest scale, which certainly constitute an appropriate conclusion from an observational viewpoint. The fact that from a traditional perspective a highly homogeneous universe would only be allowed to come into existence as a consequence of the kind of creation out of 'nothing' that would occur through inflation, therefore, no longer constitutes a decisive argument in favor of inflation theory, because from my perspective an initial period of accelerated expansion is no longer necessary to produce such an outcome.

Now, even though it should be possible for the opposite action pair creation processes which are taking place on a very short time scale to create real matter particles that do not immediately annihilate back to the vacuum when the rate of expansion is as large as it was during the first instants of the Big Bang, one cannot expect similar creation processes to spontaneously occur in the vacuum at later times, even if we assume that the densities of positive and negative matter energy that would be produced would necessarily be homogeneous, so that space may continue to expand uniformly at the same critical rate everywhere at once inside the existing universe. Indeed, while the distributions of matter energy may be required to remain homogeneous, for reasons I will explain in section 3.9, the rate of expansion itself cannot remain as large as it was during the first instants of the Big Bang, even if the densities of positive and negative energy particles which are continuously being created and annihilated in the vacuum remain unchanged as expansion proceeds, because once real matter is created as a consequence of the rapid rate of expansion, then in the context where only the matter with a given energy sign influences the rate of expansion measured by an observer with that energy sign, it follows that the rate of expansion necessarily slows down to the point where it is no longer large enough to allow the creation of more matter.

Thus, it must be clear that while it is only when we require matter to be uniformly distributed initially that it actually needs to be created out of nothing through opposite action pair creation processes during the Big Bang (for reasons I will explain later), what prevents those creation processes from persisting long after the first instants of the Big Bang is not the absence of a constraint concerning the homogeneity of the matter distribution produced

at a later epoch, but really the slowing down of the rate of expansion, which is a consequence of the very process of matter creation that took place initially and this is true even independently from how unlikely the initial conditions required for some hypothetical process of inflationary expansion to occur actually are.

In any case, it emerges that the often met remark to the effect that the observed equilibrium between open and closed universe is improbable, as it requires a delicate balance between the kinetic energy of expansion and the gravitational potential energy of matter, is irrelevant, because on the basis of the hypothesis that energy must be conserved when matter is created out of nothing during the first instants of the Big Bang, such an observation, far from being improbable, is actually unavoidable, because even if the energy contained in the gravitational field was to deviate from zero and space was to be curved on a global scale, a compensating amount of vacuum energy would immediately arise that would make space flat, so that both the total, average density of gravitational energy and that of matter and vacuum energy would remain null, as required by the constraint of relational definition of physical attributes (for reasons I explained above). The solution to the problem of flatness provided by inflation, therefore, appears to simply be unnecessary, because even when the initial density of positive energy matter is very high, the energy of the gravitational field is required to be null in a zero energy universe, which means that the universe must necessarily have a critical density of positive matter and vacuum energy and enough kinetic energy to keep expanding forever (which is also true from the viewpoint of negative energy observers, as long as the decelerating effects of a positive cosmological constant arising from a divergence of the specific rates of expansion of positive and negative energy matter can be neglected).

It would therefore appear that the idea that the initial push of inflation is necessary to explain that there is any expansion at all is incorrect, because given that high densities of positive and negative matter and radiation energy are naturally present in the vacuum in the context where pairs of opposite action particles can materialize out of nothing on a very short time scale, then expansion at a proportionately high rate becomes an absolute necessity if energy is to be null for the gravitational field independently. It should be clear, therefore, that if one must appeal to the anthropic principle it is only in order to explain the relatively small value of the cosmological constant, because the presence of an observer does not require space to be flat, even though it is certainly true that if the universe is still expanding

at the appropriate rate for life to exist it is not only because a much larger value for the average density of vacuum energy would have accelerated or decelerated the rate of expansion to such an extent that it would have become incompatible with the presence of an observer, but also because space was required to be flat by the condition of null energy that must apply to the universe as a whole.

In the context of the above description of the process of matter creation it transpires that even if bidirectional time was assumed to be continued past the initial Big Bang singularity following a hypothetical Big Bounce, the same creation out of nothing processes would have to be responsible for the existence of all the matter in the universe. Indeed, the condition that space be flat from both the viewpoint of positive energy observers and that of negative energy observers in the moments immediately following the initial singularity in the *future* direction of time, also implies that it must have been flat from both viewpoints in the moments immediately *preceding* the time at which this maximum density state is reached. Therefore, the expansion rate following the quantum bounce in the *past* direction of time would remain critical indefinitely and could initially give rise to matter creation, just as occurs in the future direction of time on our side in time of the singularity. In fact, the distribution of matter in the 'final' state which would be reached while space collapses in the *future* direction of time in that unknown portion of history taking place before the Big Bang should be identical (from a macroscopic perspective) to the matter distribution that provides the initial boundary conditions for the current one and under such conditions we can expect that most of the matter already present would return to the vacuum by being submitted to opposite action pair annihilation processes, which means that matter must indeed be created out of nothing during the first instants of our Big Bang.

To say the truth, when bidirectional time is continued past the initial singularity the appropriate initial conditions can only be obtained if there exists a constraint for the homogeneity of the distribution of matter energy in the 'final' state which would be reached forward in time in this portion of history preceding our initial singularity. If matter was not required to be homogeneously distributed in the instants immediately preceding the singularity, then we would have to conclude that a large portion of the matter could potentially have survived the state of maximum positive and negative matter and vacuum energy densities that would not have been created out of nothing at the Big Bang as a consequence of the rapid rate of expan-

sion, because a significant portion of the particles that already existed before the Big Bang would not be allowed to annihilate with an opposite action counterpart, even if there existed equal quantities of positive and negative energy particles[9]. In section 3.9 I will explain what justifies assuming that this condition (regarding the homogeneity of the distribution of positive and negative energy matter in the instants immediately preceding or following the initial singularity) must apply that requires most of the matter to exist merely as a result of the processes of opposite action pair creation that took place during the very first instants of the Big Bang.

But regardless of whether time extends past the initial singularity or not, matter creation out of nothing would occur and the magnitude of the average, initial positive and negative densities of non-gravitational energy would necessarily be that which is characteristic of quantum gravitational phenomena, which constitutes the maximum theoretically possible magnitude of energy density and therefore it is not required that matter be created at a later time by a process of reheating following a hypothetical period of inflationary expansion (which would otherwise leave the universe totally empty) in order that the Big Bang be hot, which is certainly appropriate in the context where it may no longer be required that inflation itself occurs to explain flatness. Thus, the solution to the flatness problem which is proposed here actually allows to solve the problem of matter creation. The idea that only inflation allows to explain the relatively large 'initial' density of matter, while there is no mechanism for matter creation in a more conventional Big Bang model would therefore be incorrect. In fact, the possibility for matter to be created at such an early stage also means that it is not appropriate to consider that only Higgs field particles associated with a false vacuum state were present initially, so that one of the basic hypothesis underpinning inflationary cosmology may be considered invalid.

It is remarkable, in any case, that despite our ignorance of the exact nature of the laws which apply at the Planck time, it is nevertheless possible to predict with enormous precision what the variation of the rate of expansion of the universe was when the average densities of positive and negative matter energy were maximum. It should be clear though that the existence of matter in the first instants of the Big Bang is not the byproduct of an

---

[9]It is not clear, however, if the average, initial density of vacuum energy is also required to be similar in the two portions of history and therefore it is not possible to predict with absolute certainty whether an observer would be allowed to exist in that portion of history preceding the Big Bang.

exchange of energy with the gravitational field, as was once suggested, because matter must be created in a space where the gravitational potential does not vary much locally, given that this is the actual requirement of the constraint I will later identify as being responsible for the homogeneity of the initial distribution of matter energy itself. Yet it is also true that it is as a result of an exchange of energy with the gravitational field that the positive and negative action particles created in the first instants of the Big Bang were sometimes allowed to gain a little more or a little less energy than their opposite action counterparts and in such a way give rise to small local variations in the densities of positive and negative matter energy at a level which is allowed by the above discussed constraint which limits the magnitude of initial density fluctuations.

If the preceding considerations are valid and the initial, average, specific density of positive matter and vacuum energy and the initial, average, specific density of negative energy matter[10] are both fixed very precisely to their critical value, then it would mean that the current, average, specific density of negative energy matter ρ_{mat}^{--} (measured by a negative energy observer) must be higher (in negative territory) than the average, specific density of positive energy matter ρ_{mat}^{++} (measured by a positive energy observer) by an amount equal to the current absolute value of vacuum energy density ρ_{vac}^{+} (as depicted in figure 3.3). This is because both the average, specific density of negative energy matter plus vacuum and the average, specific density of positive energy matter plus vacuum must have remained critical if they originally were, given that a flat geometry is the one configuration whose radius of curvature does not change with time.

It is not the average, specific densities of positive and negative energy matter alone which must remain critical, but really the sum of each of them with the observer independent measure of vacuum energy density. This is allowed because the same positive cosmological constant has opposite effects on the specific rates of expansion of positive and negative energy matter (measured by their own respective same-energy-sign observers) which means that the magnitude of the average, specific density of negative energy matter ρ_{mat}^{--} must indeed have decreased at a rate slower than that at which the av-

---

[10]I'm still using the expression 'positive energy matter density' to refer to the density of positive matter *energy* when the context clearly indicates that I'm not merely referring to the density of positive energy *particles* as was the case in certain portions of the above discussion.

erage, specific density of positive energy matter ρ_{mat}^{++} decreased in the course of expansion, so that ρ_{mat}^{--} must still be larger than ρ_{mat}^{++} by a value equal to the magnitude of the growing positive vacuum energy density ρ_{vac}^{+}, which is already known to be larger than the average, specific density of positive energy matter (both visible and dark). This is a simple consequence of the fact that the density of vacuum energy does not depend on the nature of the observer that measures it (when the sign of energy of matter is itself assumed to be observer independent as in a traditional context), while the total density of energy that determines the variation of the expansion rate measured by a given observer naturally adjust to remain critical by either adding positive vacuum energy to matter energy (as is the case for positive energy matter) or by subtracting this vacuum energy from matter energy (as is the case for negative energy matter).

Thus, the total average energy density ρ_{tot}^{++} that governs the expansion rate of matter from the viewpoint of a positive energy observer would currently be equal to the sum of the average, specific density of positive energy matter and the larger positive density of vacuum energy, while the total average energy density ρ_{tot}^{--} that determines the evolution of the expansion rate of matter from the viewpoint of a negative energy observer would presently be equal to the sum of the *positive* vacuum energy density and the larger (in magnitude) average, specific density of negative energy matter (positive vacuum energy is subtracted from negative matter energy). This means that the magnitude of the total, average, positive energy density ρ_{tot}^{++} is larger than the magnitude of the total, average, negative energy density ρ_{tot}^{--} despite the fact that the average, specific density of negative energy matter ρ_{mat}^{--} is larger than that of positive energy matter ρ_{mat}^{++}. Yet the specific density parameters Ω^{++} and Ω^{--} associated with those total average positive and negative energy densities are still equal to 1 given that the current expansion rate measured by a positive energy observer (through a determination of the Hubble constant H_0^+) is larger than the present rate of expansion measured by a negative energy observer (through a determination of the Hubble constant H_0^-), which is allowed to occur because the same positive cosmological constant that accelerates the expansion rate measured by a positive energy observer, contributes to further decelerate that which is measured by a negative energy observer. This can be considered to confirm the validity of the conclusion I have derived to the effect that even the homogeneously distributed portion of positive vacuum energy should exert a gravitational force on negative energy matter on the cosmological scale, unlike a homogeneous

$$-\rho_{mat}^{--} = \rho_{mat}^{++} + \rho_{vac}^{+}$$

$$\rho_{tot}^{++} = \rho_{mat}^{++} + \rho_{vac}^{+}$$
$$\rho_{tot}^{--} = \rho_{mat}^{--} + \rho_{vac}^{+}$$

$$\rho_{tot}^{++} > -\rho_{tot}^{--}$$
$$H_0^{+} > H_0^{-}$$

$$\Omega^{++} = \Omega^{--} = 1$$

$$\rho_{vac}^{+}$$

$$\rho_{mat}^{++}$$ $$-\rho_{mat}^{--}$$
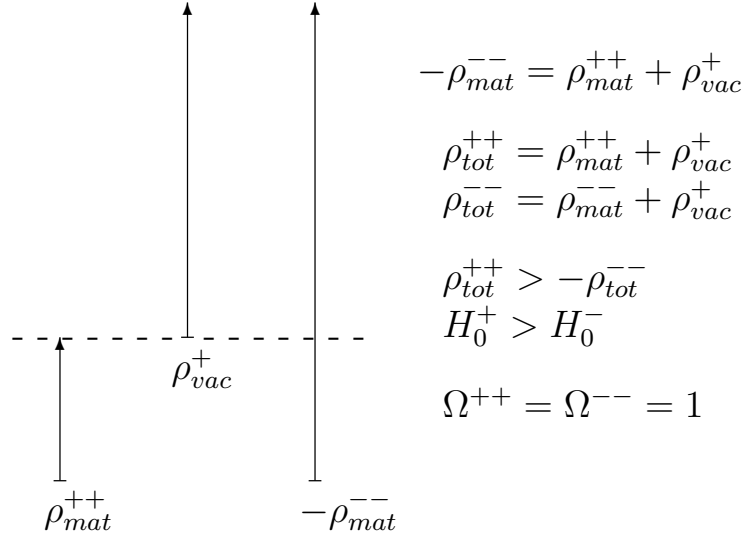
Figure 3.3: Relative magnitudes of the current, average, specific densities of positive and negative energy matter and current vacuum energy density. The magnitude of the specific density of negative energy matter ρ_{mat}^{--} is larger than that of positive energy matter ρ_{mat}^{++} by a measure equal to the positive vacuum energy density ρ_{vac}^{+}. The total energy density ρ_{tot}^{++} that governs the expansion rate experienced by a positive energy observer is larger than the magnitude of that which is experienced by its negative energy counterpart ρ_{tot}^{--} by the same measure. The current Hubble constants H_0^{+} and H_0^{-} experienced by positive and negative energy observers are modified in opposite ways by the presence of positive vacuum energy and therefore the related specific density parameters Ω^{++} and Ω^{--} remain equal to 1.

distribution of positive energy matter (both visible and dark).

In the context where the cosmological constant must be assumed to grow with time as a consequence of the diverging specific rates of expansion of positive and negative energy matter (as I have explained in section 3.2) there would be more positive vacuum energy to accelerate the rate of expansion of positive energy matter at later times, but this additional positive energy would also contribute to the total density of energy that determines the curvature of space experienced by a positive energy observer, which means that this density would remain critical if it initially was and the same is true for the total average energy density that determines the curvature of space experienced by a negative energy observer. This is allowed as a consequence of the fact that vacuum energy is conserved independently from the energy of matter and can actually be created even when it does not exist initially, because it is compensated by an associated variation of gravitational potential energy which under such conditions can actually grow (reach larger negative values) along with the expansion, exactly as would occur during a hypothetical phase of inflationary expansion.

Here one may recall the conclusion I arrived at in section 3.3 to the effect that despite the fact that, as time goes, the missing mass effect (which arises mostly from the presence of local variations in the density of vacuum energy) becomes more pronounced around visible structures in the matter distribution, the amount of positive energy dark matter (like that of negative energy dark matter) cannot be assumed to rise on the global scale. In light of the developments introduced in this section, it would appear that this conclusion is fully justified, because if the total amount of positive or negative energy matter was allowed to grow in such a way, then in the present context a contradiction would arise. Indeed, when the energy must remain null for the universe as a whole, then the average density of matter and vacuum energy are required to remain critical, so that if additional amounts of positive and negative energy dark matter were produced as a result of the growing inhomogeneity of the matter distribution, then the expansion rate would have to increase in proportion to the additional amount of matter present in a co-moving volume, despite the fact that a larger than expected positive or negative matter density would actually contribute to slow down the specific rates of expansion. One must therefore conclude that the energy which arises from the portion of missing mass effects attributable to variations in the density of vacuum energy was already present in more diffused form around elementary particles before it began to clump around large

astronomical objects composed of visible matter.

To summarize, we are in a situation where the magnitude of the sum of all positive contributions to the average, initial density of non-gravitational energy, which must be equal to the magnitude of the sum of all negative contributions of the same kind, is fixed to the maximum value that is determined by the natural scale of quantum gravitational phenomena. Under such conditions it is required that the sum of all such energies be null, given that the energy of the gravitational field must itself be null and space be flat from the viewpoint of both positive and negative energy observers in the context where a non-vanishing gravitational energy would necessarily give rise to an arbitrarily large magnitude of vacuum energy which would actually make the space flat. But the contribution of vacuum energy is also limited by the fact that the cosmological constant must not have a magnitude that would be incompatible with the emergence of an observer at later times. What allows such large matter densities as must then be present to be created out of nothing is the fact that the opposite action particle pairs which are naturally fluctuating in and out of existence on the shortest scale can avoid being submitted to immediate annihilation with other opposite action particles as a consequence of the very large initial rate of expansion that is required by the very condition of zero gravitational energy, which can only be satisfied when the kinetic energy of expansion measured by an observer with a given energy sign balances the opposite gravitational potential energy attributable to matter with the same energy sign. As a result, the average, specific densities of positive and negative energy matter are initially set to their critical value to a very high degree of precision (given that vacuum energy density must then provide a negligible contribution) and under such conditions space remains flat at all later times on the largest scale, even when vacuum energy becomes dominant.

It may perhaps appear contradictory to assume that matter needs to be created from nothing, given that I'm also suggesting that the magnitude of the energy densities which existed during the Big Bang is that which is characteristic of quantum gravitational phenomena. But in fact there is no incongruity here, because there is indeed available very large densities of positive and negative matter energy initially, only without a very rapid expansion rate the universe would have remained in a vacuum state, because this matter would have annihilated back to nothing, as it does all the time in the vacuum under ordinary circumstances. The vacuum is full of energy, only it usually does not materialize as real matter, for reasons I have explained in

section 1.9. The problem that there was with the traditional approach is that if we required a zero energy universe we could not balance the positive energy of matter in a flat universe, so that it always looked like imposing a condition of energy conservation to the Big Bang could never allow gravitational energy itself to be null, despite the fact that the kinetic energy of expansion actually is the exact opposite of the gravitational potential energy of matter for a flat universe. This is the reason why we failed to understand that applying a condition of zero energy to the universe as a whole could provide the basis for an explanation of the flatness of space that would not require assuming that the null energy of the gravitational field determined by a positive energy observer is a mere coincidence or an outcome of inflation.

3.6 The problem of time asymmetry

It is remarkable that at this point into my discussion I have already been able to provide independent solutions to two of the worst fine-tuning problems of cosmology guided merely by an unwavering confidence in the validity of well-known physical principles. It is significant also that both the solution to the cosmological constant problem and that which was proposed to the flatness problem involved considering the balancing effects of negative energy matter in order to provide additional constraints on the values of physical parameters. But before I can address other aspects of the inflation problem it will be necessary to delve a little deeper into what really constitute the many facets of the problem of time asymmetry from a classical viewpoint. This will allow me to properly identify the nature of the deep contradiction that still dwells at the heart of theoretical physics as a result of the apparent incompatibility between the time-symmetric laws of classical mechanics and particle physics and the unidirectional laws of thermodynamics and statistical mechanics.

Before engaging in a discussion of the problem of time asymmetry what one must first decide is whether irreversibility is real or whether it is a mere consequence of the way we describe the state of a system. It has been argued in effect that it is only as a consequence of adopting a particular coarse-graining and due to the choice that is made regarding what details of the microscopic state of a system are to be ignored, that irreversibility occurs. If that was the case then the continuous increase of entropy which under certain conditions appears to characterize the evolution of physical systems

with a large number of microscopic degrees of freedom would be a purely subjective notion significant merely in the context where there are practical limitations on our ability to perceive the evolution of a physical system down to its most intricate details. Under such conditions, even if entropy (as a measure of the number of possible, distinct, microscopic states of a system that are compatible with an appropriate choice of macroscopic parameters chosen to describe its evolution) was to vary, the changes which are taking place would have no fundamental significance and the observation of certain regularities regarding entropy growth would not require explanation, given that the quantity involved would merely be a subjective notion. But despite the fact that such a viewpoint is still quite popular among those who have not seriously studied the question of the origin of time asymmetry it is no longer considered by most specialists as an appropriate solution to the problem of the origin of irreversibility, but rather as a attempt at easily disposing of the problem without really explaining anything.

It was pointed out by Roger Penrose that the growth of entropy involved in most irreversible thermodynamic processes is so large that it is only marginally dependent on the choice of coarse-graining. Thus, it appears that the degree of *appropriateness* of any particular coarse-graining itself varies dramatically in the course of certain processes which are occurring all the time in our universe. The truth is that even if we were to follow the detailed evolution of all the microscopic physical parameters of a large system in a non-equilibrium state, certain aspects of this evolution could still be characterized as unidirectional. What this means is that we are not just shuffling an initially well-ordered deck of cards (to use a simple analogy) which would merely be loosing a subjective amount of structure. When we are considering an ordinary deck of cards all configurations are equivalent despite the particular significance we attach to the 'ordered' configuration. But in our universe the changes which are taking place when entropy is observed to be growing can be characterized in a more objective way, due to the nature of that portion of entropy that is attributable to the gravitational field. Indeed, the measure of entropy associated with black hole event horizons does not grow merely as a consequence of adopting a certain *arbitrary* definition regarding what parameters should characterize the macroscopic state of the system and therefore it gives rise to a less subjective notion of irreversibility. Another distinction of the evolution which is actually taking place on a macroscopic scale in our universe is that the probability to return to a former state of lower entropy never stops diminishing, because the entropy

is in principle allowed to grow without limit.

It must be clear though that this is not just a consequence of the expansion of space. It was once suggested in effect that the growth of entropy associated with all irreversible processes could be a consequence of universal expansion, given that the thermodynamic arrow of time is oriented in the same direction as what is sometimes called the cosmological arrow of time, which is merely the direction of time in which space is expanding. But it was later pointed out that this assumption is inappropriate, because in such a context it would be impossible to explain why the arrow of time should immediately reverse when space begins contracting, as would eventually occur in a universe with a negative cosmological constant (or from the viewpoint of a negative energy observer in our universe). Indeed, the expansion of space is a global phenomenon, while an expanding gas in a container is a local phenomenon which we have no reason to expect would be so drastically affected by what happens to the relative motion of distant galaxies as to start behaving anti-thermodynamically and retract into a smaller volume the moment space would begin contracting on a global scale. This conclusion is certainly appropriate given that if we were to assume that space contraction alone is sufficient to give rise to a reversal of the arrow of time then we should probably also have to assume that the thermodynamic arrow of time reverses in the presence of a strong enough, attractive, local gravitational field, while of course there is no evidence at all that this is happening.

It is usually understood, however, that while we are allowed to consider entropy as missing information, an objective characterization of temporal irreversibility does not require assuming that information is actually vanishing from reality when entropy is rising. If ignorance is growing it is only because the macroscopic parameters we use to describe the state of a system are leaving aside an increasingly larger portion of the information that would be required to accurately describe its exact microscopic state. Thus, even if certain physical parameters which allow to objectively assess the growth of entropy evolve irreversibly, the amount of structure present on a microscopic scale remains unchanged as those transformations are taking place. It is simply the fact that regardless of how well chosen they are, macroscopic parameters are increasingly less efficient at providing a full description of the structure contained in the exact microscopic state of our universe, that makes it look like information is being lost when the number of microscopic states which can potentially be occupied is growing with time.

In other words, it is merely the difficulty to keep track of all the changes

taking place in the most detailed description of the state of a system that is growing with time in an irreversible way, but no information, or no microscopic structure is really vanishing in the process. When one recognizes that there does exist a minimally coarse-grained definition of the state of a system associated with a maximum level of knowledge of its microscopic configuration, then one has no choice but to also recognize that it provides a measure of information that is unchanging. In the next section I will show how certain usually unrecognized variations in the amount of information required to describe the exact microscopic state of the gravitational field are crucially involved in allowing information to be conserved, even when the growth of entropy constitutes an objective change. But it is already possible to acknowledge that the conclusion that entropy growth does not require the minimally coarse-grained measure of information to vary is appropriate from a theoretical viewpoint, because the conservation of information is a requirement of quantum unitarity (or of Liouville's theorem in a classical context), as I have mentioned in section 2.11.

Now, if entropy is indeed increasing in the future from the viewpoint of an appropriately defined choice of coarse-graining, then it means that entropy was definitely smaller in the past. What is deduced from observations, in fact, is that entropy continuously decreases in the past, in every place we look and as far back in time as we can probe. This is a condition that is far more constraining than simply assuming that the universe is not in a state of thermal equilibrium at the present time, which would certainly also allow entropy to grow larger in the future. What we might be justified to expect in effect is that entropy should rise in the past, just as it does in the future, given that it is not already maximum at the present moment. This would appear to be implied by the fact that there is a higher probability that such states be reached as evolution takes place randomly, because there is a much, much larger number of allowed microscopic states compatible with a condition of higher entropy than there are microscopic states compatible with a condition of lower entropy. Only for an isolated system with a finite number of microscopic degrees of freedom would there be a chance that evolution could momentarily take place toward a lower entropy state as a mere statistical possibility. Such fluctuations would not constitute violations of the second law of thermodynamics given that this law is probabilistic in nature. Thus, we may consider that the evolution we observe to be taking place in general in the future direction of time is in line with expectations from both classical and statistical mechanics.

The real problem is with the past. Due to the time-symmetric nature of fundamental physical laws it would appear in effect that when a macroscopic physical system with many independent microscopic degrees of freedom evolves in the past direction of time, starting from a present non-equilibrium state of relatively low entropy, its entropy should grow (regardless of the details of its microscopic configuration) for the exact same reason that we expect its entropy to grow in the future when evolution occurs in a random way. But in our universe entropy was clearly not larger in the past than it now is and the truth is that there is no evidence from astronomical observations that any entropy decreasing phenomena has ever taken place and no written account of any person having ever observed any significant departure from constant, or continuously increasing entropy at any occasion in our entire history. Thus, while we can determine the probability of the statistically significant properties of future configurations from a knowledge of the current state of a system, the probability of past configurations cannot in general be appropriately estimated based on that same knowledge. In fact, even if entropy was continuously increasing in the past from its present non-maximum value we may still have a problem, because from the forward time viewpoint the evolution that would have taken place in the past would have occurred with diminishing entropy in the future and this aspect would also be unexplainable unless we are dealing with a *momentary* fluctuation. Thus, it seems that what must be explained is not merely why it is that entropy does not increase in the past, but why it is not already maximal and unchanging in both the past and the future.

It was suggested that the conclusion that entropy should increase in the past may not be valid, because even a macroscopic system with a large number of independent microscopic degrees of freedom could perhaps be so carefully prepared that it would be allowed to retrace an unnatural entropy decreasing evolution as it evolves backward in time. Thus, it was argued that it is the details of the present microscopic state of the universe that explains that it evolves toward apparently less probable states in the past. But unsurprisingly this argument dates back to a time when quantum chance and classical instability had not yet been discovered. In the present theoretical context, however, such an argument simply no longer makes sense, despite the fact that it is often still used to try to justify the kind of evolution that is taking place in the past direction of time. The hypothesis that a reversal of the motion of every particle in an irreversibly evolving system would bring it back to its preceding lower entropy state would actually be true only for

a very limited period of time, as short in fact as the system is large and its entropy growth in the future significant[11].

It is certainly right that a true reversal of time would actually have to involve more than a simple reversal of the motion and rotation of all components of a system, as I explained in chapter 2, but even if such a time reversal operation was applied to the whole universe there is absolutely no reason to believe that in the absence of any constraint the past evolution would have more chances of evolving toward lower entropy states, because the only violation of symmetry that might occur as a result of such a time reversal would not be such as to allow anti-thermodynamic evolution. In any case, even if we were to assume that a system could be so carefully prepared that despite the known sensibility to initial conditions which exists even in a classical deterministic context and despite the inherently random nature of quantum processes, the system would nevertheless follow an evolution so unlikely that its entropy would be continuously decreasing all the way back to the first instants of the Big Bang with absolute precision, we would still be left with having to explain why it is that the present state of the universe happens to be of such an unlikely nature that it allows this kind of awkward evolution to take place. Clearly, this attempt at explaining the occurrence of the lower entropy states into which the whole universe evolves in this direction of time we call the past cannot be considered satisfactory.

What is also problematic with the assumption that the entropy reducing evolution which we observe to take place in the past direction of time could be merely the outcome of a precise adjustment of the present microscopic state of the universe is that even if we take this as an explanation for the diminishing entropy we still cannot explain why such an adjustment does not occur for the future instead of the past, because even if that was the case it would simply seem like the past is replaced with the future and the future with the past and we would still not be able to explain why there is in

---

[11]The experiments which are sometimes mentioned as having confirmed that a reversal of the motion of all particles in the final state of a macroscopic system are observed to induce anti-thermodynamic evolution are misleading, because the processes involved take place under carefully controlled conditions, where random perturbations are absent over the totality of the short period during which the phenomena occur and therefore they merely confuse us into believing that the mystery of the continuous diminution of entropy that is taking place in the past direction of time is explainable as being the mere consequence of an improbable configuration of the present state, while this is clearly impossible under more general conditions.

effect an asymmetry. What we should actually expect to observe, if it was a precise adjustment of initial conditions that explained the occurrence of time-asymmetric behavior, is a situation where entropy would be continuously decreasing in various regions of the universe whose initial microscopic states would have been carefully prepared so as to produce anti-thermodynamic behavior, but not all in the same direction of time, that is, not all in the past direction for all locations. There is absolutely no reason to expect that such carefully prepared systems would all be set so as to evolve with diminishing entropy in only one particular direction of time, because time itself does not impose such a requirement. But we do not observe multiple oppositely directed arrows of time in our universe and this is precisely what would have to be explained for such an approach to be made valid. We cannot assume that the reason why entropy decreasing evolution is not occurring toward the future from time to time in certain locations is that the precise initial conditions required to produce it are too unlikely, while we would also be assuming that the precise 'final' conditions required to produce a decrease of entropy in the past are, for their part, allowed to occur, even if they are no less improbable. The rules of probability applied to initial conditions would lead us to predict that entropy should increase in the past just as it increases in the future and therefore they cannot alone explain the existence of an arrow of time, even if they do at least explain why it is that entropy does not decrease in the future.

Now, even if we were to recognize that the situation in which multiple coexisting subsystems would be set to evolve with decreasing entropy in opposite directions of time would probably be highly unstable, as the precise configuration required to produce a decrease of entropy in a given region would be subject to interference by what happens in another region where entropy would be decreasing in the opposite direction of time, there is no reason to believe that such a mixture of oppositely evolving subsystems should, through interference, give rise to a universe with a single well-defined direction of its arrow of time, as required by observations, that is, by our memory of past events. What must be clear is that if we do not expect to frequently observe such carefully prepared subsystems evolving with diminishing entropy in the future, then we should not expect to observe the entire universe itself to evolve in such a unnatural way in the past, but this is precisely what is happening all the time and if that is indeed the case then there must be another explanation to it.

It is only as a consequence of the fact that, for practical reasons, our

thought processes are always functioning in the direction of time in which entropy is rising (thereby giving rise to a psychological arrow of time) that we usually fail to recognize that the kind of evolution that takes place in the past direction of time is amazingly abnormal from a purely probabilistic viewpoint. Thus, while it is certainly true that the present state of the universe is relatively unlikely configured, for example in the sense that if time was reversed a local tendency for matter particles to disperse would momentarily turn into one for particles to convene, while a tendency for wavefronts to spread would turn into one for wavefronts to converge, this is explainable as merely being a consequence of the fact that the original state in the past that gave rise to the present 'final' state was itself in a highly unlikely configuration, even from a purely macroscopic viewpoint. It's not the final states which are inexplicably organized, but really the initial state (in the distant past) that gave rise to them.

One of the oldest attempt at solving the problem of the origin of the arrow of time which must also be considered inadequate was originally proposed by Ludwig Boltzmann, the originator of the kinetic theory of gases. It was based on the recognition that there always occur fluctuations to lower entropy states for randomly evolving isolated systems which are in a state of thermal equilibrium. On a very long time-scale it should sometimes happen that those fluctuations would be so significant as to bring even a system in thermal equilibrium into a state with an entropy so low that any subsequent evolution would likely be characterized by a continuous increase of entropy. Thus, it was proposed that the universe, as the ultimate isolated system, really starts in a maximum entropy state, which would presumably be a likely state to be randomly chosen as our initial conditions, and then remains in such a state during most of its existence, but that once in a while, as it evolves in either the past or the future, it simply fluctuates to a much lower entropy state from which it would naturally be expected to evolve with continuously increasing entropy back to its more likely maximum entropy state in the same arbitrarily determined direction of time, which we would then call the future regardless of its actual (relative) orientation. The fact that such an evolution would perhaps appear to be similar to that which we presently observe to occur at the level of the universe as a whole then suggests that this is what explains the continuous growth of entropy in one single direction of time that characterizes the evolution of all systems which have not yet reached back a state of thermal equilibrium.

It should be clear, however, that in such a context the only reason we

would have to expect to observe the universe in a phase of continuously growing entropy instead of finding it in one of the much, much more common phases of unchanging maximum entropy would be that this entropy growth is necessary for the presence of an observer which can witness such an evolution. Indeed, the fact that we are allowed to experience a memory of past events and to have a persistent conscious existence is dependent on the condition that there exists a well-defined thermodynamic arrow of time. The problem, however, is that if such a requirement was to be satisfied merely as a consequence of the occurrence of a fluctuation in an otherwise unchanging maximum entropy state, then we should not expect to observe entropy to be so low in all parts of the universe and as far back in time as the epoch of the Big Bang. A much more localized and ephemeral fluctuation that would provide the observer with no records of a long lasting history would do just as well for allowing such a condition at the present time and given that such a fluctuation would be more likely to occur than a long-lived fluctuation involving the entire universe, then based on this kind of argument what we should experience is a short-lived fluctuation.

The question, therefore, remains: Why is the universe evolving irreversibly in one single direction of time in all locations and throughout its entire lifetime? One cannot hope to satisfy the requirement imposed by the fundamental time-symmetric physical laws by simply postulating that the universe actually evolves without any constraint either in the past or the future, because that would leave the very property of irreversibility unexplained. As Boltzmann himself appears to have realized, the entropy fluctuating universe scenario is ineffective for explaining this very constraining aspect of reality and therefore cannot count as a valid solution to the problem of time asymmetry.

Now, the fact that I'm suggesting that the random nature of elementary physical processes and the sensibility to initial conditions is what allows to reject the possibility that it could be a precise adjustment of the present conditions that would completely explain the diminution of entropy that is observed to take place in the past direction of time does not mean that I'm adhering to the opinion that irreversibility is occurring at a fundamental and irreducible level in our description of physical processes, as was once proposed by some of those who pioneered the study of chaotic systems. I do not believe that we must equate unpredictability and randomness with irreducible time asymmetry, even if in its most general form statistical mechanics, as a probabilistic theory, is dealing with systems in non-equilibrium states whose

evolution is inherently irreversible. The fact that quantum field theory can be considered to be a more fundamental instance of statistical mechanics, while it is definitely and explicitly a time-symmetric theory, clearly indicates that my position is justified. It would certainly not be appropriate to abdicate the requirement of symmetry under a reversal of the direction of time simply to provide an explanation for the observed unidirectionality of thermodynamic processes in the context where our most valuable physical theories are all time-symmetric at the most elementary level of description.

The difficulty that we are experiencing in trying to identify the constraint that allows to derive irreversible evolution from time-symmetric physical laws should not be allowed to become a justification for abandoning some of the requirements we have very good reasons to believe must constitute part of a fully satisfactory solution. We would not be wise in rejecting a theoretical framework that works so well, even if it may seem that it cannot explain every aspect of reality, simply to follow an alternative approach which also cannot be made to describe all significant aspects of reality. The challenge consists in actually explaining irreversibility, not in decreeing that it is the foundation of reality when this would require abandoning most of everything else we have learned. I believe that the fact that we have not yet been able to achieve this objective is not an indication that our most fundamental theories are wrong, but merely a proof that we still do not fully understand all the consequences of the physical principles upon which they were built.

It is important to note in this regard that it has also been proposed that it is perhaps a fundamental irreversibility of the quantum measurement process that allows to explain the asymmetry of the evolution in time of observable physical phenomena that does not appear to characterize the evolution that takes place in between measurements. But while I do not want to enter into a discussion of how irreversibility intertwine with quantum theory, I must point out that it would be circular reasoning to assume that it is the measurement process that gives rise to thermodynamic irreversibility, while it is already recognized that it is the irreversibility of the processes taking place in the environment with which a quantum system interacts that is involved in giving rise to the decoherence effect that characterizes all quantum measurements. But even if we were to follow such a route it is not clear what would explain that this same unidirectionality does not instead operate toward the past rather than the future. After all, there is no sign of an intrinsic asymmetry regarding the direction of time in the equations of quantum theory. Why would quantum evolution always pick the same one particular direction

of time instead of another during those processes that can be qualified as measurements? Once again, even if for pure convenience it was assumed to be the case that quantum theory or a hypothetical process of actualization of potentialities was to show preference for one direction of time instead of another, we still would be left with as great a mystery to explain, because time itself does not provide the means for such a distinctive feature to arise.

I do agree that irreversibility (just like time itself) is real and constitutes an objective aspect of physical reality and is not just a consequence of some arbitrary choice regarding the level of coarse-graining, but what I will try to demonstrate is that the suggestion that it is no longer appropriate to conceive of reality in terms of elementary particles obeying time-reversible physical laws is not justified, even when we are dealing with complex systems which exhibit strong non-linearity or highly irreversible evolution. As we will progress, it will become clear that the idea that there should be no laws more fundamental than those which currently apply only under those particular conditions is excessive in proportion to the very specific nature of that most extraordinary property of physical reality we are trying to explain.

3.7   Gravitational entropy

Now that I have properly defined and circumscribed the problem of time asymmetry I would like to explain what the reasons are that allow me to believe that an objective notion of entropy growth may exist despite the fact that the amount of information needed to completely describe the microscopic structure of a system is required to be rigorously conserved. This will constitute an important development, because it is ultimately the non-subjective character of that portion of entropy variation which is attributable to the gravitational field that enables one to conceive of the irreversibility that characterizes the evolution of certain macroscopic physical systems as being an objective property even under conditions where gravitation does not appear to be involved, because, as I will emphasize in the following section, all the entropy growth that is taking place in our universe must ultimately be attributed to the initial conditions of low gravitational entropy that existed in the remote past. As a consequence of the progress I have achieved in better understanding the properties of gravitational entropy I will also be able to provide a decisive solution to the problem of the violation of the conservation of information which appears to take place in the context where the expan-

sion of space is continuously creating new elementary quantum gravitational units of space in the vacuum.

What is already known concerning gravitational entropy is that it grows when the mass of an astronomical object and the strength of its gravitational field are rising. Thus, when gravitational attraction is involved, the natural tendency for matter to spontaneously disperse into a larger volume of space is overcome and the decreasing entropy of matter that follows is compensated by the even larger increase of entropy presumably attributable to the gravitational field. In fact, we currently have no exact definition for the entropy attributable to the gravitational field in a general context and it is merely a knowledge of the exact formula for black hole entropy that allows one to estimate the magnitude of this entropy in the absence of event horizons. In any case, the prevailing character of gravitational entropy means that when a large enough amount of matter is present in a given volume of space, particles with the same sign of energy are allowed to become more densely packed, because such an evolution is favored from a thermodynamic viewpoint in the context where there are more possible microscopic configurations of the gravitational field compatible with a state of higher density. Only the expansion of space could eventually allow this natural tendency to be surmounted if the growth of entropy attributable to expansion becomes rapid enough that it overcompensates the growth of entropy that occurs through the formation of inhomogeneities.

What I would like to point out is that the presence of event horizons provides us with a unique set of macroscopic physical parameters which allow a natural definition of coarse-graining and therefore an objective measure of entropy growth. Indeed, what emerges from the semi-classical theory of black hole thermodynamics is that the area of a black hole event horizon provides an exact quantitative measure of entropy that is not subject to any arbitrary choices concerning the coarse-graining. Thus, any change to entropy which is reflected in a variation of the mass or the surface area of a black hole constitutes a non-subjective change which cannot be attributed merely to the choice of macroscopic parameters, as under such conditions no other macroscopic parameter is available to define an alternative measure of entropy. Yet the information loss which might appear to be taking place when a black hole absorbs low entropy matter cannot be considered real, because as I mentioned in section 2.11 it seems that the information about the microscopic state of the matter that was submitted to gravitational collapse is encoded in binary form in the microscopic degrees of freedom associated with

the elementary units of area on the event horizon of the object and is released when the black hole decays through the emission of thermal radiation.

It is certainly true that when the exact evolution of a system that is not in a state of thermal equilibrium cannot be followed down to its most intricate details we may lose sight of information concerning its exact microscopic state and therefore more information than is available afterward may be needed to describe it. But even in the context where entropy growth is an objective change (due to the non-subjective nature of the definition of coarse-graining that is made possible by the existence of black hole event horizons) this information loss merely constitutes a practical limitation and in principle the information necessary to describe the exact microscopic state of a system submitted to such irreversible changes could still be retrieved. What must be understood is that the growing amount of missing information that would be required to completely specify the state of all the microscopic, binary degrees of freedom on the surface of a black hole of increasing mass also constitute an objective change that reflects the existence of a growing amount of microscopic structure in the gravitational field that is not dependent on any arbitrary choices regarding the coarse-graining.

Thus, the objectivity of the measure of missing information associated with the area of a black hole event horizon does not only mean that information is not lost when such an object absorbs low entropy matter, it also implies that the growth in the amount of information required to describe the microscopic state of the gravitational field constitutes an objective change. But what's even more significant is that when one recognizes the appropriateness of the assumptions that allowed me to derive an exact measure for the entropy of elementary black holes based on a knowledge of the relevant discrete variables that characterize the fundamental states of matter under the most extreme conditions, then it becomes necessary to admit that new degrees of freedom, which characterize the exact microscopic state of the gravitational field, are being created when the entropy of a black hole is growing, because the amount of missing information which would be required to specify the exact microscopic state of all the matter particles which were captured by the gravitational field of a black hole is not large enough to account for its entropy growth.

What my findings from section 2.11 regarding the existence of a relationship between black hole entropy and the degrees of freedom associated with the discrete symmetry operations indicate, in effect, is that the amount of missing information which would be required to completely specify the micro-

scopic state of matter particles is actually decreasing when matter is captured by the gravitational field of a black hole. In fact, part of the additional information required to describe the microscopic state of matter in the absence of a macroscopic event horizon would ultimately need to be attributed to the existence of additional microscopic structure in the gravitational field, given that this missing information merely allows to specify the value of physical parameters such as the energy signs of the elementary black holes which are present in the vacuum (as I explained in section 2.11) and therefore the amount of missing information contained in the microscopic state of matter itself is not really being reduced when the density of matter is rising locally. But given that, according to my analysis, the amount of missing matter information is certainly not growing either when the mass of a black hole is rising, while the total amount of missing information (the entropy) is growing faster than the mass of the object (which rises in proportion to its matter content), then one has no choice but to recognize that the amount of missing information which would be required to describe the microscopic state of the gravitational field is indeed rising when local gravitational fields grow stronger, at least in the presence of a macroscopic event horizon.

It should be clear, therefore, that when matter assembles into a macroscopic black hole, the number of microscopic degrees of freedom associated with the gravitational field grows larger, even while the number of microscopic degrees of freedom associated with matter particles is being reduced as a result of the constraints exerted on their states of motion by the gravitational field of the object. Thus, while information about the exact microscopic state of the matter that fell into a black hole is not provided by the macroscopic physical parameters that describe the object and may therefore appear to be lost, an even larger amount of information is created at the same time which actually contributes to increase the entropy of the black hole and which arises from the emergence of new microscopic degrees of freedom associated with the gravitational field itself. This means that the amount of missing information that would be necessary to completely describe the exact, unknown, microscopic state of the gravitational field appears to be growing faster than we would expect if information was conserved and it is also this additional increase and not only the progression of our ignorance concerning the intricate details of the microscopic state of matter that is responsible for the growth of entropy that occurs when a black hole absorbs matter.

Now, given that I will later argue that the growth of inhomogeneities in

the matter distribution, which is the source of stronger gravitational fields, provides the dominant contribution to entropy increase in our universe (because the entropy of matter itself does not change much as a consequence of expansion) then it would appear that irreversibility actually arises mostly as a consequence of the growth of gravitational entropy. What is crucial to understand under such conditions is that this irreversible evolution as well as the growth in the amount of missing information which is associated with it cannot be considered subjective features of reality, precisely because they can be associated with the presence of event horizons which constitute natural boundaries enabling a unique definition of coarse-graining that is entirely determined by the strength of local gravitational fields.

It must, in effect, be recognized that an objectively larger amount of information is required to describe in complete detail the structure that emerges in the gravitational field when inhomogeneities develop in the matter distribution (or in the distribution of dark matter that is attributable to local variations of vacuum energy). I believe that it is merely because we do not benefit from the guidance of a fully developed quantum theory of gravitation that we haven't yet realized that the amount of missing information is actually growing faster than would appear to be allowed, when a gravitational field gains in strength as a consequence of a local increase in the energy density of matter (we often hear about people claiming that information may be lost when matter falls into a black hole, but I have never heard anyone complaining about the growth of missing gravitational information). One has no choice, however, but to recognize that when a gravitational field gains in strength, even in the absence of a macroscopic event horizon, an objective increase in the amount of missing information, which is not due merely to increased ignorance, and a truly significant growth of entropy, which is not dependent on any subjective definition, are taking place.

It would, therefore, appear that a concept of temporal irreversibility can actually be defined which is not dependent on any arbitrary choices regarding which macroscopic parameters are significant for the description of physical systems with a large number of independent microscopic degrees of freedom. Indeed, given that the values taken by the macroscopic physical parameters associated with the event horizon of a black hole (from which the measure of its entropy is determined) are not dependent on any arbitrary choices (regarding the coarse-graining) it follows that irreversibility can be characterized as an objective aspect of the evolution that is taking place in our universe whenever the gravitational interaction is involved.

What one would normally object concerning this characterization of gravitational entropy is that the growth of missing information which can be expected to occur when stronger gravitational fields develop, would then appear to violate the constraint of conservation of information that is imposed by quantum theory. Yet, I have also come to understand that despite the fact that, locally, the amount of missing information may actually be growing faster than would appear to be allowed as a mere consequence of growing ignorance concerning the microscopic state of matter, the total amount of information required to describe the exact microscopic state of our universe does not really change when gravitational fields gain in strength and therefore the requirement of conservation of information is not violated. What must be clear in any case is that either information is always conserved, or else it never is, and given that the latter conclusion does not appear to be valid from a fundamental viewpoint then it must be recognized that the additional information which appears to be produced when a black hole forms already existed before it contributed to the measure of gravitational entropy associated with such an object, just like the information contributed by matter itself. But if a larger than allowed change in the amount of missing information is indeed impossible, then it means that any such variation would need to be compensated somehow.

Indeed, what implies that the additional growth in the amount of missing information which is associated with stronger local gravitational fields can be objectively characterized is merely the fact that it occurs as a result of adopting the natural definition of coarse-graining that is provided by the measure of spacetime curvature associated with the presence of macroscopic event horizons as natural boundaries with well-defined macroscopic physical parameters. If those considerations are appropriate, however, then it becomes necessary to recognize that the growth in the amount of missing information that is taking place as a consequence of a local increase in the density of matter can only be compensated by a change in the amount of information which would itself be independent from any particular definition regarding the choice of coarse-graining and therefore we can already expect that it would be attributable to additional changes in the strength of local gravitational fields.

The situation we face, therefore, is one in which the amount of information that is missing (and which determines the coarse-grained measure of entropy) is continuously rising, even though it is only in situations where stronger gravitational fields develop, due to a local increase in the density of

matter (of positive or negative energy sign), that this variation can be characterized as a non-subjective change attributable in part to a real growth in the number of microscopic degrees of freedom, rather than to our growing ignorance of the details of the exact microscopic state of the matter which is the source of those gravitational fields. Once we recognize that under such conditions the amount of missing information required to completely describe the microscopic state of the gravitational field is indeed growing faster than would appear to be allowed, then what remains to explain is how information can nevertheless be conserved, as would presumably be required in a quantum gravitational context. In fact, what allows me to conclude that the amount of missing information is growing faster than would appear possible when stronger gravitational fields develop as a result of the formation of a matter overdensity is not merely the results of my analysis of the nature of the microscopic degrees of freedom of matter constrained by the gravitational field of a black hole, but the very fact that it also appears necessary to assume that there is an opposite variation of the same kind that occurs when gravitational fields grow stronger as a result of the formation of an *underdensity* in the large scale matter distribution, which suggests that it is only as a consequence of the fact that there arises a compensation between those two variations that the measure of information can be left invariant regardless of how fast it varies locally.

What I'm suggesting more exactly is that given that a higher than average matter density appears to be associated with an additional amount of missing information which was not present initially, due to the fact that a larger amount of information is required to completely describe the detailed microscopic state of the gravitational field under such conditions, then it should necessarily be the case that a correspondingly smaller amount of information would be required in order to completely describe, with the same level of precision, the microscopic state of the gravitational field associated with a lower than average matter density. You may recall that in section 1.6 I explained that a void in the cosmic distribution of positive energy matter must actually be considered to exert a gravitational repulsion on the surrounding positive energy matter due to the fact that the presence of such a void implies an *absence* of gravitational attraction which would otherwise compensate that which is attributable to the surrounding matter distribution, whose center of mass is always located in the exact position of the particle experiencing its gravitational influence. But if those gravitational forces are in effect attributable to an absence of gravitational interaction with the pos-

itive energy matter that is missing in the void, then it means that a lesser amount of information would be required to describe the microscopic state of the gravitational field as a result of the presence of such a void.

In the context where the initial matter distribution may have been very uniform to begin with, this conclusion would imply that any additional increase in the amount of missing information necessary to describe the microscopic state of the gravitational field attributable to the formation of a local matter overdensity would be compensated by an exactly corresponding decrease of information attributable to the presence of the underdensity that must necessarily form in the surroundings of this overdense structure in order to allow it to grow. As a result, I can deduce that despite the objective nature of the local changes that take place in the measure of information when the matter distribution is growing more inhomogeneous, information is always rigorously conserved, even when this evolution involves an alteration of the macroscopic parameters associated with event horizons. But it must be clear that those conclusions only apply in situations where it is gravitation that provides the dominant contribution to entropy change and where it is a variation of information associated with the gravitational field that compensates another variation in the amount of missing information required to describe the microscopic state of the same field, because it is only under such conditions that we can expect objectively defined changes in the amount of information to occur locally.

Thus, when the density of matter grows larger than its average value there is an increase in the amount of missing information that is attributable merely to growing ignorance concerning the exact microscopic state of the matter, but there is also an additional increase in the amount of missing information that is attributable to the creation of new degrees of freedom in the microscopic state of the gravitational field. When the density of matter becomes smaller than its average value, however, there occurs a corresponding *decrease* in the amount of information that is attributable to the elimination of certain degrees of freedom which originally existed in the microscopic state of the gravitational field despite the uniformity of the initial matter distribution. It is this decrease, attributable to the formation of an underdensity in the uniform matter distribution, that compensates the additional unaccounted increase in the amount of missing information attributable to the formation of the corresponding overdensity and which would otherwise violate the condition of conservation of information. In other words, when the mass of an astronomical object is growing, more information than would

appear to exist initially is required to describe the exact microscopic state of its gravitational field, whose higher strength is responsible for most of the entropy growth that occurs under such conditions. But in an originally smooth matter distribution the growth of mass in one place can only arise when a corresponding diminution of mass takes place in the surrounding area and from a certain perspective less information would appear to be required to describe the exact microscopic state of the gravitational field when the matter density is reduced below its average cosmic value in such a way, even if a stronger gravitational field would seem to be produced locally as a result of such a change.

What is happening, therefore, is that given that it is not necessary to describe the microscopic degrees of freedom which are absent in the gravitational field as a result of the absence of gravitational interaction with the matter that is missing, it follows that the microscopic state of the gravitational field can be completely specified using a smaller amount of information. In fact, as I will explain below, it is this dependence of the amount of microscopic structure on the strength of *attractive* gravitational forces that allows one to understand why it is that when the density of matter grows *larger* in a local region of space, more information than may appear to have existed initially is required to describe the microscopic configuration of the gravitational field, because there is no a priori motive for assuming such an outcome, despite the fact that it appears to be required by the semi-classical theory of black hole thermodynamics in the context of my account of the constraints applying on the microscopic state of matter particles reaching a future singularity. What is crucial to understand, however, is that a local decrease in the amount of information necessary to describe the microscopic state of the gravitational field does not necessarily translate in a reduction of gravitational entropy.

Indeed, if the density of matter is only allowed to decrease in a given region of space when a compensating increase takes place in its vicinity, it follows that the information loss that occurs as a result of the formation of an underdensity in the matter distribution only serves to increase the amount of missing information necessary to describe the exact state of the gravitational field associated with the creation of the corresponding matter overdensity. But this means that information which was available before the change took place, as a consequence of being associated with microscopic states of the gravitational field which were not constrained by the presence of a macroscopic gravitational field or event horizon, would now have to be accounted

for as missing information that merely contributes to raise the gravitational entropy of the overdensity whose mass grew at the expense of the formation of the underdense structure. Therefore, even if the amount of information necessary to describe the exact microscopic state of the gravitational field is diminishing locally, the total measure of gravitational entropy would still be growing in the universe, as any measure of entropy does under ordinary conditions when information becomes unavailable, despite the fact that it does not vanish from reality.

This property of gravitational entropy to rise globally at the expense of a local decrease in the measure of available information contained in the same force field is reflected in the fact that the strength of local gravitational fields is actually growing, even when the stronger fields are attributable to an absence of gravitational interaction consequent to the formation of a void in the matter distribution. As a consequence, the changes occurring when a void is forming in the matter distribution are still likely to take place when gravitation is predominant, because under such conditions they are actually favored from a thermodynamic viewpoint. It remains, however, that the gravitational fields attributable to the presence of voids in the positive energy matter distribution do not have the exact same thermodynamic properties as the similar gravitational fields attributable to the presence of negative energy matter overdensities, as will be emphasized below.

Anyhow, once it is recognized that the amount of information required to describe the microscopic state of the gravitational field must be reduced when the density of matter diminishes below its average value locally, then it becomes possible to conclude that the total measure of information concerning the microscopic state of the gravitational field always remains constant globally, as required by quantum theory and despite the objective nature of the growth of entropy that occurs when gravitational fields gain in strength as a result of the development of inhomogeneities in the matter distribution. Thus, even though the variations in the amount of information can be characterized in a more objective way when gravitation is involved, there is no fundamental difference between those changes and the ones that take place when there is no significant variation in the strength of local gravitational fields[12]. An objectively defined diminution of information does occur

---

[12]In fact, it may be that it is always the case that when a large enough static force field develops, the growth in the amount of missing information which occurs as a consequence of the creation of additional microscopic degrees of freedom could be qualified in an objective way and would therefore require a compensating contribution to occur in the environment.

when the density of positive energy matter diminishes below its average cosmic value and the strength of the *repulsive* gravitational fields experienced by positive energy matter grows locally and this is what allows to compensate the additional growth in the amount of missing information that occurs when a gravitational collapse is taking place in the positive energy matter distribution.

What we fail to recognize from a conventional viewpoint is not only that a *local* increase in the density of matter and a stronger attractive gravitational field give rise to an objective increase in the amount of missing information required to describe the microscopic state of the gravitational field that would appear to be larger than allowed by the conservation of information. We also fail to recognize that a *local* diminution of matter density below its cosmic average would actually give rise to a diminution in the amount of information required to specify the microscopic state of the gravitational field (given that the gravitational field attributable to such an underdensity in the matter distribution would arise from a local absence of gravitational interaction). The fact that such a compensation is required to take place if information is to be conserved can be considered to provide confirmation of the appropriateness of the results I derived in section 1.6 to the effect that, not only must voids in a uniform matter distribution be the source of repulsive gravitational fields, but that that those gravitational fields actually originate from uncompensated gravitational attraction by the surrounding matter distribution.

The conclusion that a local decrease in the density of matter must give rise to a local diminution in the amount of information required to describe the microscopic state of the gravitational field is much more profound and significant than one may perhaps expect. Indeed, despite the fact that there is a certain equivalence between the gravitational field produced by the presence of an overdensity of *negative* energy matter and the gravitational field attributable to an underdensity in the *positive* energy matter distribution, a clear distinction must nevertheless exist between those two situations with regards to thermodynamic properties. In section 2.12 I explained in effect that a negative energy black hole in a vacuum must be considered to radiate negative energy particles and to have negative temperature. Thus, if a void

---

This is perhaps a desirable hypothesis given that according to my analysis of black hole entropy (discussed in section 2.11) the fields associated with other long range interactions can actually be expected to carry their own specific measures of entropy.

in the positive energy matter distribution was deep enough over a sufficiently large region to produce a gravitational field equivalent to that of a negative energy black hole it would appear necessary to assume that it has negative temperature, given that its surface gravitational field is opposite that of a positive energy black hole and similar to that of a negative energy black hole. But temperature merely defines the relationship between energy and entropy as is well-known from Clausius' definition of entropy change through the formula $dS = dQ/T$ (where dQ is the amount of heat absorbed or released by a system with temperature T in the small time interval during which it evolves between two equilibrium states). Under such conditions if a negative energy black hole has negative temperature it must actually radiate negative energy particles, or negative heat (just like a positive energy black hole must radiate positive energy particles), so that its surface area and its entropy can diminish in the process. One might, therefore, be tempted to assume that the thermodynamic properties of a sufficiently large void in the positive energy matter distribution would be identical to those of a negative energy black hole and that such a structure would radiate negative energy particles. But that is not the case.

First of all, it must be clear that there is nothing wrong with the idea that the temperature associated with the thermal radiation of a negative energy black hole is negative. Once it is understood that this radiation process arises as a consequence of the thermodynamic requirement that local energy differences be smoothed out, even in the presence of event horizons, then it is clear that a negative energy black hole must lose negative energy if its mass is to decrease in the process. Given that positive energy matter cannot cross a negative energy black hole's event horizon and remain inside such an object, then this loss of negative energy can only occur through the emission of negative energy particles outside the event horizon. A negative energy black hole would, therefore, *release* negative heat in its environment (which constitutes a positive change for the energy of such an object) and in the process necessarily reduce its surface area and its entropy, which therefore requires the temperature of the object to be negative. But it is precisely here that the distinction between a negative energy black hole and a sufficiently deep void in the positive energy matter distribution would arise, because the thermodynamic tendency to reach equilibrium would not produce the same outcome in the case of the void in a positive energy matter distribution, despite the similarity of the gravitational fields associated with both kinds of configuration.

Indeed, while the gravitational field produced by a sufficiently deep void in the positive energy matter distribution must be equivalent to that of a negative energy black hole from an external viewpoint, in the case of the void the uniformity of the distribution of energy cannot be reestablished through the emission of negative energy particles by the void, because there is no way that such a radiative process could allow the void to regain the lost positive energy that gave rise to its growth, even if negative energy particles were present inside the structure and could surmount the growing gravitational attraction exerted on them as they would stray from the center of mass of the void. What would happen, therefore, is that equilibrium would be reached through the *absorption* of positive energy particles from the surrounding matter distribution, which is not forbidden as it would for a negative energy black hole, because the strength of the repulsive gravitational field actually decreases as a positive energy particle approaches the center of the structure, given that the equivalent mass of the object is not all concentrated in a central singularity, as is the case for an ordinary negative energy black hole. Thus, even if the temperature of a sufficiently large void in the positive energy matter distribution is negative, the structure would not be expected to reach equilibrium through the emission of negative heat, but rather through the absorption of positive heat, which would actually allow the gravitational entropy of the structure to be reduced in the process as the positive energy that is absorbed would come to replenish the void at the expense of a decrease in the density of surrounding matter overdensities.

This conclusion is actually a mere reflection of the fact that the temperature of such a void in the positive energy matter distribution, like that of a negative energy black hole, must be assumed to be negative. In section 2.12 I have explained in effect that it is when an increase of energy produces a local decrease of entropy that the temperature of a system must be considered negative. From the preceding discussion it should be clear that while a negative energy black hole satisfies this condition as a consequence of the fact that removing negative energy from it reduces its surface area and therefore its entropy, a void in the positive energy matter distribution would satisfy the same condition merely as a result of the fact that adding positive energy to it (through a reduction in the energy of surrounding positive energy matter overdensities) would also give rise to a local diminution of gravitational entropy. Thus, even if the surface gravitational field and the temperature of a sufficiently large void in the positive energy matter distribution could actually be identical to those of a negative energy black hole,

one must conclude that the thermodynamic properties of those two kinds of matter inhomogeneities are not exactly the same.

In any case it would appear that the conclusion that entropy rises when a void is forming in the matter distribution is inevitable. But if the assumption that the information required to describe the microscopic state of the gravitational field itself decreases locally under the same conditions is to be considered valid it must be further justified from a more elementary perspective. I will now explain what justifies my conclusion that a lower matter density is to be associated with a reduced amount of information in the gravitational field. What must be clear once again is that despite the apparent similarity between voids in a matter distribution and overdensities of opposite energy sign, there nevertheless exists a fundamental difference between those two categories of objects which arises from the fact that negative energy objects do not consist of voids in a positive energy matter distribution, but are rather equivalent to voids in the positive energy portion of the *vacuum*, as I emphasized in section 1.8. It must be clear also that the conclusion that the formation of a void in the matter distribution would give rise to a negative change in the amount of information required to describe the microscopic state of the gravitational field is also valid in the case of a negative energy matter distribution. The sign of changes occurring in the objective measure of information is not dependent on the sign of energy of the matter which is the source of the varying gravitational fields. Thus, the formation of a void in the negative energy matter distribution would also give rise to a local reduction in the amount of information required to describe the exact state of the gravitational field. But based on the above discussed argument, it is also necessary to conclude that a sufficiently deep void in the negative energy matter distribution should have a positive temperature just like a positive energy black hole, because its surface gravitational field is equivalent to that of such an object and this means that as it gains negative energy through absorption of negative energy radiation, a local decrease of gravitational entropy would occur, just as would be the case for a decaying void in the positive energy matter distribution.

I believe that what explains that the formation of a void in the uniform positive energy matter distribution would give rise to a negative change in the amount of information concerning the microscopic state of the gravitational field, while the formation of a void of similar magnitude in the positive energy portion of the vacuum, which can be assimilated with the formation of a negative energy matter overdensity, would produce a positive change in the

measure of missing information concerning the gravitational field, is the fact that in the absence of local variations of vacuum energy density (which can be associated with the presence of dark matter), the distribution of vacuum energy is really uniform on all scales, while the 'homogeneous' distribution of matter (visible or dark) in which a void may be produced is not really uniform on a microscopic scale. Indeed, in a uniform distribution of vacuum energy there are no persistent density variations, such as those which would be associated with the presence of real particles, and removing energy from such a perfectly uniform distribution cannot be assumed to reduce the amount of structure that would initially have been present on a microscopic scale in the gravitational field which is attributable to the presence of this energy. This is unlike the situation we have when we are dealing with what would normally be considered a homogeneous matter distribution in which there actually exist smaller scale variations in the density of energy, which create local gravitational fields which may not be apparent from a global viewpoint, but which can be as strong as the density of matter is high.

Thus, when we locally reduce the density of matter particles in a *macroscopically* uniform matter distribution we reduce the strength of the *microscopic* gravitational fields which are present in this matter distribution as a result of its own small scale inhomogeneity. But with those gravitational fields was associated a certain microscopic structure and this can only mean that in such a case we need less information to describe the exact microscopic configuration of the gravitational field, because we actually reduce the amount of structure that previously existed in this field as a result of the inhomogeneity of the microscopic distribution of matter particles. By contrast, when we increase the density of matter with opposite energy sign we produce stronger microscopic gravitational fields that were not present beforehand in the vacuum and it is only appropriate that in such a case the amount of missing information associated with the microscopic structure of the gravitational field is actually growing. This is all a consequence of the fact that more densely packed particles exert stronger attractive gravitational forces on each other, so that a reduction in the number of particles present in a given volume of space reduces the strength of the local gravitational fields which would otherwise be present on a smaller scale, while removing energy from the *vacuum* can actually generate additional variations in the microscopic state of the gravitational field, given that it is equivalent to increasing the number of matter particles of opposite energy sign.

Unlike a local reduction in the density of positive vacuum energy, a lo-

cal reduction in the average density of positive energy matter gives rise to
a diminution in the amount of information necessary to describe the micro-
scopic state of the gravitational field and this is reflected in the fact that
an underdensity in the positive energy matter distribution does not have
the exact same thermodynamic properties as an overdensity in the negative
energy matter distribution, despite the similarity of the gravitational fields
produced by the presence of both kinds of astronomical structures from an
external viewpoint. In such a context it becomes possible to actually explain
not only why it is that the amount of information contained in the gravi-
tational field must diminish when a void forms in the uniform, large scale
matter distribution, but also why it is that the amount of missing informa-
tion about the microscopic state of the gravitational field is actually growing
when the density of matter is increasing locally.

This argument concerning the distinction between local diminutions in
vacuum energy density and local diminutions in matter density would also
justify assuming that even the gravitational field attributable to an appar-
ently smooth matter distribution would contribute a certain measure of infor-
mation, despite the fact it is traditionally assumed that only the gravitational
fields associated with the presence of macroscopic inhomogeneities in the dis-
tribution of matter energy contain information. Indeed, if locally reducing
the density of matter produces a decrease of information in the gravitational
field, then it would seem that even on a global scale a certain amount of
information should be contained in the gravitational field produced by the
uniformly distributed matter, which would be reduced as a result of expan-
sion. This reduction would occur because a global decrease in matter density
would reduce the strength of the microscopic gravitational fields between in-
dividual matter particles, which would otherwise contain a larger amount of
information, just as is the case for macroscopic gravitational fields, only here
we are dealing with additional degrees of freedom which are normally left out
of a classical description of the gravitational field attributable to a uniform
distribution of matter energy. In fact, the same condition of conservation of
information which imposes a compensation between the local variations of
the different measures of gravitational field information attributable to the
formation of matter inhomogeneities appears to require that a certain mea-
sure of information be associated with the microscopic gravitational fields
which are present in a homogeneous matter distribution.

Indeed, as expansion takes place, the density of matter decreases, which
means that a reduction in gravitational field information would be attributa-

ble to this expansion and would need to be compensated by an equivalent increase in the amount of missing information. Now, it has already been proposed that the expansion of space should perhaps be considered to produce an increase in the amount of missing information, given that it would appear to continuously produce additional elementary units of space on the quantum gravitational scale, in apparent violation of the theoretical requirement regarding the conservation of information. I believe that this suggestion is valid, because according to the developments introduced in section 2.11 it appears that a larger number of elementary units of space would imply the existence of a larger number of fluctuating elementary black holes in the vacuum and a complete determination of the microscopic state of the virtual particles submitted to the gravitational fields of those objects would require additional binary units of information. But unlike those who previously discussed the issue I do not believe that this growth in the amount of missing information (which is actually a growth of missing gravitational field information) constitutes a serious difficulty, because I know that this change is compensated by the diminution of gravitational field information that is attributable to the diminishing matter density that takes place as a consequence of expansion. For this to be a valid proposal, however, it must be recognized that a variation of the average density of vacuum energy, or the cosmological constant, would not contribute to alter the total amount of information contained in the microscopic state of the gravitational field, despite the fact that, like the average matter distribution itself, the uniform portion of vacuum energy would provide a variable contribution to the gravitational field that influences the rate of expansion.

The distinction which, according to the above analysis, would exist between the variation of gravitational field information arising from a local decrease in the density of matter and the variation of gravitational field information produced by a similar decrease in *vacuum* energy density would therefore appear to constitute an essential requirement if information is to be conserved on a cosmological scale. Indeed, in the absence of such a distinction gravitational field information would vary as a result of changes occurring in the average density of both matter energy and vacuum energy and this would be problematic, because the variation of gravitational field information associated with the changes in average energy density which are occurring over the entire lifetime of the universe could not be compensated by the variation of information attributable to the growing volume of space. This would be a consequence of the fact that only the average energy den-

sity of matter varies along with the scale factor and could be reduced to a minimum value when the volume of space has become arbitrarily large, or be taken back to a larger value when the volume of space is reduced to a smaller value. The average density of vacuum energy, on the other hand, would either become dominant (regardless of its energy sign) in a universe of ever growing size and diminishing matter density, thereby giving rise to both a growing total energy density and an arbitrarily large volume of space with no possible compensation of the changes taking place in the total amount of information (as would occur from the viewpoint of a positive energy observer measuring a positive cosmological constant), or else become dominant in a recollapsing universe with growing matter density and a diminishing volume of space (as would occur from the viewpoint of a negative energy observer also measuring a positive cosmological constant), thereby again precluding the initial value of information from being conserved as it must be, that is, independently for positive and negative energy observers which experience different measures of the gravitational field.

However, once it is recognized that changes in the gravitational field attributable to variations in the *average* density of vacuum energy do not contribute any changes to gravitational field information (given the absence of persistent microscopic inhomogeneities in the distribution of vacuum energy) and therefore need not be taken into account in balancing the rising amount of missing information associated with the growing volume of space produced by expansion, then those difficulties no longer exist. From the viewpoint of a positive energy observer the average density of matter (including the equivalent density of dark matter attributable to local variations of vacuum energy density) would in effect be continuously decreasing as a consequence of the expansion of space, along with the associated measure of information required to specify the microscopic state of the gravitational field which was originally maximum (this is allowed by the fact that a reduction of negative energy matter density also contributes to lower the measure of gravitational field information despite the opposite sign of the variation of energy density itself). But at the same time the amount of missing information associated with the number of elementary units of space present within a co-moving volume (or more accurately the number of elementary units of area on the two-dimensional boundary of the same volume) would grow to some arbitrarily large value, thereby compensating the change to gravitational field information that is associated with the diminishing matter density and allowing the entropy of the universe to keep growing, while

a similar compensation would need to occur in the case of the recollapsing space eventually experienced by a negative energy observer.

But this is a valid conclusion only when the variable average density of vacuum energy does not contribute any change to the amount of information necessary to describe the microscopic state of the gravitational field[13]. I believe that this is the strongest argument which can be formulated to the effect that it is appropriate to consider that even the diminution of average matter density which is taking place on a global scale actually gives rise to a decrease in the amount of information contained in the microscopic state of the gravitational field, because it clearly implies that a variation of matter density occurring in an *macroscopically* uniform distribution of matter energy must in effect be assumed to produce changes in the microscopic gravitational fields which are different from those we would expect to occur as a consequence of a variation in the average density of the *locally* uniform distribution of vacuum energy. If I have properly conveyed the nature of the insights which have allowed me to arrive at such a conclusion then it should be clear that there is no longer a problem with the fact that the expansion of space appears to produce information. From my viewpoint, even if this growth in the amount of missing information must indeed be considered real, it would not give rise to a net increase in the total amount of information required to completely specify the microscopic state of both physical space and the gravitational field for the universe as a whole.

3.8 The initial singularity

What emerges from the preceding reflexion concerning the character of gravitational entropy is that while the amount of missing information required to describe the microscopic state of the gravitational field is growing in those places where matter is becoming more densely packed, an equal amount of information is being lost at the same time in the gravitational field as a consequence of the resulting diminution of matter density which is taking place in

---

[13]If those conclusions are appropriate it would mean that the idea proposed by certain authors that the size of the elementary units of space determined by the natural scale of quantum gravitational phenomena is perhaps itself growing with time, so that the amount of missing information associated with the total volume of space would be constant despite expansion, which should eventually give rise to a 'Big Snap' that would rip everything apart, can be considered unnecessary and this is certainly appropriate given that no such an event seems to be occurring.

the surrounding space and both changes are independent from any arbitrary choices regarding the coarse-graining. Yet, given that the information that is gained is accounted for as missing information, while the information that is lost was in principle available (as it was associated with a structure that was not constrained by the presence of a macroscopic gravitational field or event horizon), then gravitational entropy must nevertheless be assumed to rise whenever the matter distribution is becoming more inhomogeneous. What I will now explain is how significant this conclusion actually is in the context where the initial distribution of matter energy at the Big Bang appears to have been one of inexplicably high uniformity. Thus, I will argue that for what regards irreversibility, it is the measure of gravitational entropy that constitutes the significant difference between the state that emerged from the past Big Bang singularity and the state into which our universe will evolve in the far future (independently from whether it continues to expand or collapses back on itself). This discussion will set the stage for the more significant developments which will be introduced in the next section and which will provide the actual explanation for the existence of the thermodynamic arrow of time as a cosmological phenomenon.

It is important to note, first of all, that there is no paradox associated with the fact that the universe still evolves irreversibly while the initial state at the Big Bang was already one of near perfect thermal equilibrium, because as Roger Penrose first pointed out [39], under such conditions it is only the portion of entropy which excludes the contribution of local gravitational fields that is maximum. In fact, what transpires from the developments introduced in the previous section concerning gravitational entropy is that it is precisely the smoothness of the initial distribution of matter energy (which is reflected in the uniformity of the temperature of the cosmic microwave background) that is responsible for having allowed the universe to evolve irreversibly at later times, because the growth of gravitational entropy has been by far the dominant contribution to irreversible change in our universe since the epoch of decoupling. What really needs to be explained, therefore, is not why the universe evolves irreversibly despite the initial state of thermal equilibrium, but why the energy of matter was actually so homogeneously distributed initially that gravitational entropy was almost perfectly null, even if that would appear to be a highly unlikely configuration to begin with in the context where a much larger number of possibilities exist for the microscopic state of matter and its gravitational field which would not be characterized by such a uniform matter distribution and an absence of primordial black

holes.

In order to explain those facts one needs to identify the nature of the constraint imposed by the fundamental, time-symmetric physical laws on the boundary conditions at the Big Bang that is responsible for the very high level of homogeneity and the very low gravitational entropy that characterizes this initial state. We must, therefore, once again transcend our common reluctance to apply the known principles of physics to the Big Bang if we are to avoid having to modify the laws themselves in order to achieve greater overall consistency. It would be incorrect to assume that proposing a solution to the problem of the origin of time asymmetry that relies on the application of certain constraints to the initial conditions at the Big Bang would be akin to requiring divine intervention. The most fundamental principles must be assumed to be valid under absolutely all conditions, including those that existed during the Big Bang. I believe that it is our failure to acknowledge the importance of this requirement that explains most of the difficulties we currently face in theoretical cosmology.

But before we can achieve some real progress in understanding why irreversibility occurs we must first recognize that the source of most changes to entropy that take place after the epoch of decoupling and the emission of the cosmic microwave background radiation is actually to be found in the growing strength of local gravitational fields. It is as a consequence of gravitational attraction that the stars, in particular, can form and are allowed to release their radiation and it is also due to gravitation that black holes, as the objects with the highest entropy density, can form and grow more massive at the expense of a local reduction of matter density in their environment, which is also the source of stronger gravitational fields. But one need not assume that this is due to the 'fact' that gravitation is always attractive, as all that is required is that it be attractive among particles with the same sign of energy, which allows gravitational energy and therefore also gravitational entropy to be proportional to the square of the mass of an object instead of being merely proportional to its mass, as does ordinary entropy under different circumstances. In such a context it seems that a much larger number of initial states would be characterized by the presence of an abundance of black holes and other density fluctuations, while those initial conditions would not have had as much potential for allowing irreversible evolution. However, given that the presence of primordial black holes would have disturbed the process of structure formation in the initial matter distribution in ways which would have had observable consequences at the present epoch,

then it seems necessary to assume that the initial Big Bang state was virtually free of black holes and therefore it remains to explain why the universe was in such an unlikely configuration at the Big Bang.

One thing that should be clear is that the weakness of the gravitational interaction in comparison with other forces and the fact that it appears to have come into effect much later than those other interactions during the Big Bang does not mean that no constraint that would be imposed on the magnitude of local gravitational fields could be involved in determining the early conditions which are responsible for the existence of the thermodynamic arrow of time. Indeed, as I pointed out when I discussed the flatness problem, it is the gravitational interaction which is responsible for having fixed the rate of expansion itself as a function of the density of matter in the very first instants of the Big Bang and therefore it must certainly have exerted a significant influence even at the earliest epoch. Also, the fact that gravitation began to produce local gravitational collapse at a relatively late time is due precisely to the fact that the initial distribution of matter energy was so uniform to begin with and this is a constraint which is actually imposed on the magnitude of local gravitational fields (if it wasn't imposed expansion would not even have been allowed to persist locally, as I explained in section 3.5) and it would certainly be inappropriate to assume that a constraint on the initial magnitude of local gravitational fields would not have much impact as a consequence of the very fact that the magnitude of local gravitational fields was in effect so small initially.

In section 3.5 I have also explained why we can actually expect the universe to be expanding. But the fact that we are not instead observing it to be contracting at the present moment can only be explained as being the consequence of another fact, which is that the magnitude of local gravitational fields is decreasing continuously in this direction of time relative to which the universe is contracting. If we perceive the universe to be expanding it is simply because, as thermodynamic processes, our memories are formed only in the direction of time in which the inhomogeneity of the matter distribution is growing, while if the strength of local gravitational fields and the measure of gravitational entropy were growing in the direction of time relative to which the universe is contracting, then we would necessarily perceive the universe to be contracting. This is actually all that can be meant when we say that we experience the universe to be expanding, because in fact we do also 'observe' space to be contracting, but merely in the sense that we also have knowledge of the contraction of space that occurs in the past direction of time, as

we may witness by watching a backward running movie of the same events. Thus, what explains that the universe is observed to be expanding (what explains that the cosmological arrow of time is oriented in the same direction as the thermodynamic arrow of time) is the fact that gravitational entropy is practically null in the primordial Big Bang state, while it is allowed to grow to arbitrarily large values at later times and this means that if we want to explain why it is that we observe an expanding universe then we must first explain why it is that its initial state was characterized by such a low gravitational entropy.

But in the context where we must acknowledge the presence of negative energy matter in our universe, the fact that the density of matter was much larger in the past does not make the initial smoothness of the matter distribution more unexpected, as one may be tempted to assume. Indeed, even in a universe that would never have been through a maximum density state, an initial configuration characterized by a greater uniformity of the distribution of matter energy would not necessarily be more likely as a randomly chosen boundary condition for the universe, because even a diluted matter distribution could still feature inhomogeneities on the largest scale, as a result of the fact that negative energy matter can be concentrated in regions of space distinct from those occupied by positive energy matter, even if the average density of both types of matter is negligibly small. Given that it cannot even be assumed that such a large scale polarization of the two matter distributions would have no observable consequences, it is certainly not appropriate to consider that this possibility can be ignored from an empirical viewpoint. Thus, if the density had not been higher in the past, the most likely configuration for the distribution of matter energy might still have been one of higher inhomogeneity, because there usually exist more microscopic configurations of matter and its gravitational field for which positive and negative energy matter are not mixed up in a perfectly smooth manner.

In any case, the hot Big Bang did occur and the distribution of matter and radiation energy was in effect homogeneous to an inexplicably high degree initially. If this hadn't been the case then macroscopic event horizons would abound in the primordial state and even if the magnitude of density fluctuations was not large enough to prevent the universe from expanding in most locations, what we would observe would be a much different world. The problem, therefore, is that regardless of its volume in the far past it seems that the observable universe should have begun its evolution in a state where the energy of matter would already be highly inhomogeneously distributed

and strong local gravitational fields would be present with which would be associated an arbitrarily large measure of entropy. But if the initial state was not of such a nature then it means that something must have constrained the universe to have a much lower gravitational entropy initially, because this does not appear to be a natural configuration to begin with when all possibilities are allowed.

It should be clear, however, that the simple fact that the universe must be expanding locally if an observer is to be present to witness an absence of inhomogeneities does not provide strong enough a constraint to explain that the initial distribution of matter energy was as smooth as it is observed to be, even if the presence of event horizons would indeed prevent space from expanding locally. The energy of matter could be much more inhomogeneously distributed than it currently is and expansion would still be allowed to proceed unaffected in most locations, even if a large number of primordial black holes had been present initially. It is merely the fact that the inhomogeneity is not as pronounced as it *could* have been that is unexplainable.

What must be understood is that the homogeneity of the initial distribution of matter energy does not arise merely from the low magnitude of local variations in the energy of elementary particles (which can be compensated by *local* variations in gravitational energy, as I explained in section 3.5), but must also be apparent in the near absence of large scale disparities in the distribution of positive and negative energy matter particles, which actually allows most of the matter to be produced as opposite action particle pairs during the Big Bang. From the viewpoint of a negative energy observer this particularity is especially significant given that the only difference which will exist, apart from the larger magnitude of vacuum energy density, between the expanding and the recollapsing phases of the universe's history has to do with the fact that in the recollapsing phase the dissociation of the positive and negative energy matter distributions will actually be much more pronounced, as a result of the gradual polarization of the matter distribution along energy sign which can be expected to occur in the context where particles with the same sign of energy are submitted to mutual gravitational attraction, while concentrations of matter with opposite energy signs gravitationally repel one another.

For this discussion to be meaningful, however, one must also understand that there are strong motives for believing that even in the presence of negative energy matter it is still appropriate to consider that there arises a state in the past which from a classical viewpoint would be characterized as con-

sisting in a spacetime singularity. Indeed, what should be clear based on the developments introduced in section 1.6 is that a homogeneous distribution of negative energy matter would not exert an influence on the rate of expansion of positive energy matter and would not diminish the strength of the gravitational field attributable to the presence of this matter, despite the fact that negative energy matter would in general exert a repulsive gravitational force on positive energy matter. This conclusion follows from the description of negative energy matter as being equivalent to the presence of voids in the positive energy portion of the vacuum and the acknowledgement that the void of cosmic proportion that must be associated with a homogeneous distribution of negative energy matter cannot give rise to uncompensated gravitational attraction from a surrounding distribution of positive vacuum energy, which would otherwise be the source of the gravitational repulsion that would arise from the presence of such a void.

Once this is recognized, it becomes possible to predict that if the initial matter distribution is sufficiently homogeneous on the largest scale, then nothing can prevent the formation of the trapped surface which according to classical theorems would give rise to a past singularity, even if one of the axioms of the theorems is that matter must always have positive energy. It is only the inappropriateness of the traditional description of negative energy matter as being the source of absolutely repulsive gravitational fields that makes it seem like the presence of such matter could prevent the formation of a past singularity (or the occurrence of a state of maximum matter energy density, as one would rather need to assume in a quantum gravitational context).

It would therefore appear that the very uniformity of the matter distribution which is responsible for giving rise to the existence of a thermodynamic arrow of time is actually required in order that the existence of a past singularity, or an initial state of maximum matter density, be considered unavoidable. This is a decisive observation whose significance will be made more explicit in the following section. But if there really is a singularity at the onset of time, then given that I have already been able to identify the fundamental degrees of freedom which characterize the state of matter that collapsed into an ordinary black hole singularity based on a semi-classical description of this singular state, it would seem that it is possible to characterize the initial state from which our Big Bang originated with much better accuracy than is usually assumed. I must emphasize, therefore, that what I have in mind when I'm referring to an initial singularity is not a state where the laws of

physics would actually break down, but simply a state where the average, positive and negative densities of matter and vacuum energy have reached the maximum theoretical values of energy density determined by the natural vacuum-stress-energy tensors which enter the generalized gravitational field equations introduced in section 1.15.

In any case, if we are to assume that there must in effect be a singularity, or a state of maximum matter density at the beginning of time then it seems necessary to assume that this singularity is also different in certain respects from an ordinary black hole singularity. First of all, even if the initial state that emerged from the past singularity at the Big Bang had been highly inhomogeneous it would not be expected to have given rise to the same evolution as that into which a future Big Crunch singularity would go from a backward in time viewpoint, because whereas the state that would emerge from a Big Crunch singularity would evolve back to a more homogeneous state, the state emerging from an initial Big Bang singularity with the same level of matter energy inhomogeneity would not evolve toward a more homogeneous state, because in our universe future evolution is unconstrained. Therefore, a highly inhomogeneous distribution of matter energy emerging from a past singularity may not evolve at all from a thermodynamic viewpoint, as it would already be in one of its most likely maximum gravitational entropy states and only the expansion of space could perhaps allow some irreversible evolution to take place by allowing the positive and negative energy black holes initially present to slowly decay in the ever growing volume of space. As a result, no reversed gravitational collapse or white hole phenomenon would occur that would release objects of lower entropy and only thermal radiation would be emitted, without much significant changes actually taking place, in perfect accordance with the second law of thermodynamics.

But while the Big Bang is not the time-reverse of a Big Crunch or of a black hole gravitational collapse it also appears that the initial singularity is different from a future singularity owing to the fact that it does not give rise to an initial state characterized by large fluctuations in matter energy density with which would be associated a very large gravitational entropy, such as would be the case for the final state of a generic future singularity. This observation makes it even more apparent that what is occurring in the past direction of time in our universe is not what one would expect to happen as a mere consequence of the contraction of space. The initial singularity was of such a nature that it could not constitute the outcome of a gravitational collapse of the kind that would occur in a universe in which local gravitational

fields are growing in strength. The universe is changing as it collapses in the past direction of time, but not in the way one would expect in the absence of a constraint that operates a continuous decrease in the magnitude of the inhomogeneities present in the distribution of matter energy.

What's significant as well is that the presence of past singularities appears to be restricted to the one known initial singularity from which the Big Bang emerged, even if there does exist solutions of the gravitational field equations that would appear to describe processes which would be the time-reverses of a black hole gravitational collapse. All the evidence indicates that the hypothetical white hole processes which could be described using those solutions never occur in our universe. I believe that if those solutions do not represent processes that can be observed in the forward direction of time in our universe it is because they would allow gravitational entropy to decrease in this direction of time, even if such an evolution is thermodynamically unlikely in the absence of a specific constraint. Indeed, white holes would expel low entropy matter at an arbitrarily high rate, which would reduce their masses and the area of their event horizons faster than would be allowed as a consequence of the emission of thermal radiation (this has nothing to do with negative energy black holes expelling positive energy matter) so that the processes would involve a decrease of gravitational entropy in the future. It should be clear, therefore, that the Big Bang does not constitute a generic white hole, even though it originates from a past singularity.

The only motive one might have to assume that generic white holes could exist would be that in all likeliness the gravitational entropy of a black hole should rise in the past just like it does in the future, so that from the forward in time viewpoint the evolution taking place during the same period of time would actually appear as a fluctuation involving a decrease of gravitational entropy that would persist until the present moment is reached. But the problem is that even if the present state may be compatible with the occurrence of such a phenomenon, there appears to be something that constrain evolution in the past direction of time to take place with continuously decreasing gravitational entropy despite the apparent improbability of this evolution, and this is precisely what remains unexplained. If white holes are never observed, therefore, it is simply because such processes would require a decrease of gravitational entropy in the future (which is unlikely) or equivalently a continuous increase of gravitational entropy in the past direction of time (which for some reason appears to be forbidden). Therefore, if we can understand why the state that emerged from the initial Big Bang singu-

larity had minimum gravitational entropy, then we may also be allowed to understand why there is only one such past singularity.

At this point it should be clear that even though black holes are the objects associated with the highest possible density of gravitational entropy, it would not make sense to simply assume that the most likely initial state for the universe would be one for which all matter would be contained in one giant black hole, because even a closed universe with a highly homogeneous distribution of matter energy could be considered to satisfy this condition. What is required for gravitational entropy to be maximum is that matter energy be as inhomogeneously distributed as possible even while the universe is in the process of collapsing into a higher density state from a backward in time viewpoint. The relevance of this remark is made more obvious when we are considering a universe that contains both positive and negative energy matter. Indeed, in such a context, the state with the highest gravitational entropy would necessarily be one for which the distributions of positive and negative energy matter would be completely dissociated in such a way that all the matter would be contained in opposite energy black holes with arbitrarily large masses whose magnitude would be limited solely by the amount of matter in the universe and the time available for the inhomogeneities to form (if they are not already present to begin with). What must be understood, therefore, is that there is no a priori motive for assuming that a high level of dissociation between positive and negative energy matter could not also apply to the initial Big Bang state (regardless of the fact that its matter density is maximum) if such a configuration is in effect favored from a thermodynamic viewpoint, because a universe that would evolve without constraint as space is contracting in the past direction of time would have more chance to reach such a configuration, not merely despite gravitational repulsion, but as a result of it.

Now, it was once suggested that the smoothness of the initial distribution of matter energy might only be apparent and that a state of higher inhomogeneity might have existed initially that was later made uniform through various smoothing processes. But given that such processes would have released a large amount of heat that would have modified the temperature of the cosmic microwave background to an extent that appears to be incompatible with measurements, then it appears that even if the smoothing could occur at the appropriate time and on the appropriate scale, its outcome would not agree with observational constraints. Furthermore, if the distribution of matter energy had been highly inhomogeneous before any such process could

smoothen it out, the magnitude of those inhomogeneities would have rapidly been amplified under the effect of the gravitational interaction and it would have become even more difficult to give rise to the homogeneous distribution that is revealed by measurements of the temperature of cosmic microwave background radiation. Indeed, the same argument implies that the initial state cannot have been *perfectly* uniform, otherwise the universe could not have evolved into its present state early enough to allow for the existence of stars, galaxies and other large scale structures, which means that the constraint responsible for the high level of homogeneity of the initial state must not be so restrictive that it would imply a complete absence of energy fluctuations.

What constitutes the most significant difficulty for the smoothing hypothesis, however, is the fact that the existence of cosmic horizons would have forbidden any such process from ironing out inhomogeneities above the scale determined by the size of the horizon at the time when the CMB was released and therefore we should not observe uniformity on the largest scale if the homogeneity of the distribution of matter energy is attributable to smoothing processes obeying the requirement of local causality. An intrinsic limit is actually imposed on such processes that would prevent them from producing the kind of homogeneous state which emerged from the Big Bang and therefore it appears appropriate to conclude that regardless of any other difficulty, conventional smoothing processes should probably not be considered a viable explanation for the homogeneity of the initial distribution of matter energy.

As a consequence of the clear inadequacy of conventional smoothing processes and in the absence of a better alternative it is still widely believed that inflation may be the cause of the very high homogeneity of the universe's distribution of matter energy which is reflected in the small amplitude of cosmic microwave background temperature fluctuations. However, I think that the occurrence of this hypothetical process of accelerated expansion would not be of much help in explaining the observed time asymmetry that characterizes cosmic evolution, because there is no reason to expect that a contracting universe would evolve toward a more homogeneous configuration during the epoch that would precede a hypothetical phase of exponentially accelerated *contraction* which would then take the universe back to a more likely state of maximum inhomogeneity. If inflation could perhaps explain why the universe evolves in an otherwise unnatural way (from the viewpoint of the growth of gravitational entropy) between the moment when matter emerges from the

initial singularity and the instant at which inflation ceases, it could not explain why it evolves toward greater homogeneity from far in the future and back toward the time before the universe would presumably begin to contract at an exponentially accelerated rate into the initial singularity, now with naturally growing inhomogeneity.

Even if inflation may give rise to a homogeneous universe forward in time, a Big Crunch would not be expected to occur with decreasing inhomogeneity forward in time, unless the state immediately preceding the exponentially accelerated contraction into the final singularity would be required to be as smooth as the state which was produced in the past following ordinary inflation. But assuming that this would occur would amount to require that causality operates backward in time from the final singularity, instead of forward in time, because from the viewpoint where causality operates from the past to the future, a Big Crunch would be more likely to occur with increasing inhomogeneity in the future right up to the moment when inflation would perhaps take place in reverse and merely increase the inhomogeneity that would already exist even further and produce an inhomogeneous final singularity. Assuming that this is not what occurs would amount to postulate without motive that causality must rather operate backward from the instant at which matter emerges from the future Big Crunch singularity and until the moment when the universe would begin recollapsing after having reached its maximum volume, so that the period of inflation that would occur backward in time from the point at which matter emerges from the final singularity would give rise to a homogeneous state *after* inflation in the past direction of time. But there is no a priori reason not to assume instead that it is a highly inhomogeneous final state existing *before* the phase of exponentially accelerated contraction that gives rise to the inhomogeneous state that would occur in the future direction of time following this phase of exponentially accelerated contraction, as we may expect based on the hypothesis that causality still operates forward in time.

The problem is that the hypothesis that classical causality operates forward in time from the past singularity is necessary for the conclusion that inflation would necessarily produce a homogeneous state, because if it was assumed that it is the events in the future that can influence what occurs backward in time until the moment when matter would start contracting at an exponentially accelerated rate back into the initial singularity, then the state we would expect to obtain following inflation, from the forward in time viewpoint, would still be a state of maximum inhomogeneity, while this

does not correspond to reality. What must be understood is that even if we simply interchange future and past we are still facing a mystery, because if it is the future that influences the past and if inflation operates backward in time so as to smooth out the state emerging from a future Big Crunch singularity, instead of giving rise to a homogeneous state forward in time beginning from the inhomogeneous state that emerged from the past singularity, then we simply reverse the direction in which irreversible evolution would take place and we still have no explanation for why causality in effect operates in this particular direction of time (which we would then call the future) and not in the opposite one, while this is precisely what we are trying to explain. Indeed, classical causality, or the rule that past events always have an influence on future events and not the opposite is simply a manifestation of irreversibility or time asymmetry and if this property is assumed to characterize our universe without question, then it cannot be used to explain irreversibility itself. Therefore, assuming that inflation necessarily produces a highly homogeneous state from a more likely inhomogeneous state amounts to assume without justification the very outcome we want to derive, which means that inflation is not valid as an explanation of the origin of time asymmetry that would arise from the necessity of a homogeneous initial state (following inflation).

Thus, it is not merely the fact that the highly unlikely initial conditions necessary for an exponentially accelerated phase of contraction may not exist in the state preceding the final future singularity of a collapsing universe (while they are assumed to have existed when matter emerged from the initial past singularity) that would make inflation ineffective in predicting irreversible evolution, as certain authors suggested. The fact that inflation itself requires quite unique initial conditions to occur does not even need to be taken into account to conclude that such scenarios do not really allow to explain the observed asymmetry of the evolution of gravitational entropy. To actually explain the unlikely homogeneous state that emerged from inflation during the Big Bang using the hypothesis of inflation itself we would have to predict that this process operates in both the future and the past directions of time to produce a homogeneous state out of the generic inhomogeneous initial states that would emerge from both the initial and the final singularity and this would require that the direction of time relative to which inhomogeneities are growing mysteriously reverses when the universe starts contracting, when its volume would be maximum (or at any arbitrarily chosen intermediary time indeed) and as I previously explained there is absolutely no reason to

expect that a reversal of the thermodynamic arrow of time associated with the growth of gravitational entropy would occur when space would begin contracting on a global scale.

At this point it is necessary to mention that a variation of the more conventional attempt at explaining cosmological time asymmetry by making use of inflation theory which was proposed more recently postulates that it is through the process of creation out of 'nothing' that symmetry with respect to the direction of time can be reintroduced in our description of cosmic history. In this context the process of creation out of nothing is actually a process of creation out of a preexisting, extended empty space and therefore I would not myself describe it as creation out of truly nothing. Anyhow, what is proposed is that the initial state of our universe is actually a fluctuating vacuum state with minimum positive energy matter density, which we may perhaps consider to be a likely state from a thermodynamic viewpoint given that it is associated with an arbitrarily large number of elementary quantum gravitational units of space which could be the outcome of a prior phase of expansion that would have taken place in an open universe. Of course this is not the state in which observations indicate our universe began, but it might be possible to assume that what happened is that the universe emerged out of a local fluctuation in this extended vacuum and that it is inflation that is responsible for having allowed the high density state so produced to start expanding at a critical rate and if this is indeed the case then the universe could perhaps be considered to necessarily begin in a state (preceding inflation) that is not so unlikely from the viewpoint of gravitational entropy, even if this would otherwise be unexpected. The idea is that this kind of process could reproduce in the future as expansion again gives rise to an extended vacuum state and perhaps to a universe whose future is not that different from its past. While such a model would not solve any of the difficulties I previously identified as affecting an explanation of time asymmetry dependent on inflation, it could at least allow us to take comfort in having obtained a description of our universe's history that would allow it to begin in what may appear to be a more likely initial state (the extended vacuum state preceding inflation).

The problem I have with this description, however, is not merely that in the context where the existence of negative energy matter constitutes a necessary assumption, an unchanging, extended vacuum state characterized by an absence of local gravitational fields may not be the most likely configuration for our final state (which would also be required to constitute our initial

state). Neither does it have to do mostly with the fact that the proposed solution depends on the contribution of a hypothetical process like inflation, which is now known to itself require highly unlikely initial conditions of a distinct nature if it is to actually give rise to the appropriate outcome. The more unavoidable difficulty has to do with the fact that as a tentative explanation of time asymmetry it would suffer from the same reliance shared by more traditional approaches on the implicit assumption that there is already a favored direction of time.

Indeed, despite what is usually assumed, an extended vacuum state in our past could only arise out of a prior phase of expansion that would occur in the future direction of time. If a large volume of space is to remain nearly empty for a sufficiently long time that fluctuations in the vacuum are perhaps allowed to give rise to the creation of an entire universe, then this space must have been expanding prior to the event and this expansion can only take place in one direction of time at once. Indeed, I will explain in the next section that it is not possible to simply assume the existence of an expanding low density universe without assuming that it has emerged out of a state of maximum matter density created out of truly nothing at some point in the past (even if bidirectional time could be extended past such an event), which would still require that the unlikely distribution of matter energy and the low gravitational entropy that characterize this earlier initial state actually be explained. Thus, it transpires that any such model would merely involve indefinitely postponing the problem of explaining the apparently unlikely nature of the initial singularity by requiring the existence of yet another singularity that would exist at an earlier epoch and that would still need to be responsible for the observed time asymmetry.

If we believe that the initial conditions at the Big Bang must be subjected to the same constraint of likeliness as applies to the configurations of matter which are reached through random evolution under more general circumstances, then the fact that it does not appear that this initial state could have been produced by chance alone means that there must be an explanation for this anomaly, but this explanation cannot be found in the traditionally favored cosmological models based on inflation theory. It must be noted again that the anthropic principle would be of no use in trying to achieve such a goal, because if the initial conditions are freely determined they would not be required to be so highly constrained as they appear to have been when matter emerged from its maximum density state. Indeed, a state as homogeneous as that which appears to have existed in the remote

past is so unlikely to have arisen randomly that even the chance occurrence of an observer in a universe with a less thermodynamically favorable initial state would be a more likely phenomenon in comparison. If the universe was initially characterized by such a low gravitational entropy it is because it necessarily had to go through such a constrained initial state at least once in its lifetime. What I will now explain is why this conclusion should have been expected all along.

3.9   The horizon problem and irreversibility

So here we are, having actually ruled out the possibility that the high degree of homogeneity of the matter distribution in the primordial universe could be due to any conventional or inflationary smoothing processes, but with apparently no option left to explain this remarkable fact. Although this outcome may be quite perplexing, the attentive reader may already have perceived a glimmer of light on the cosmic horizon. Indeed, when one carefully looks at all those failed attempts I believe that one cannot avoid getting the feeling that it is the very fact that there exists a state of maximum matter density in the remote past that must constitute the basis of a consistent explanation of the origin of the anti-thermodynamic evolution that is taking place in the past direction of time in our universe and which is giving rise to an ever more homogeneous matter distribution. Indeed, the presence of the Big Bang is the decisive aspect that differentiate the remote past from our far future in the context where we can no longer expect a Big Crunch to occur. What I will now explain is that there is actually a requirement for the density of matter to be maximum at a certain point in the history of the universe that does not just follow from the fact that space must expand or contract and this actually allows to explain why it is that the universe did not came into existence in an extended vacuum state with negligible matter density. But, quite remarkably, this same requirement is also responsible for having produced a maximum density state so exceptionally configured that it guarantees that all future evolution will take place irreversibly.

Before embarking on an explanation of how it can be that the initial Big Bang state allowed the emergence of an arrow of time, however, I would like to first recall my earlier discussion from section 3.5 concerning the nature of the processes of matter creation out of nothing that gave rise to our entire universe. There I mentioned that energy conservation alone does not require

that the energy of the matter which is created out of nothing be uniformly distributed, as even in a zero energy universe *local* fluctuations in matter energy density can be compensated by a non-zero energy of the gravitational field that is significant for both positive and negative energy matter. But if the energy of the matter created out of nothing had been as inhomogeneously distributed as it can be, macroscopic black hole event horizons would abound in the early universe.

One might be tempted to argue that the uniformity of the initial distribution of matter energy arises as a consequence of the fact that the presence of very large density fluctuations in the initial Big Bang state would not allow matter to be produced out of nothing by processes of opposite action pair creation, in the context where those processes can only occur on the scale of distance characteristic of quantum gravitational phenomena. Indeed, it appears that opposite action particles could not be created by pair out of nothing if they are already concentrated in black holes with opposite energy signs initially. But given that local fluctuations in matter energy density could be present, even in a zero energy universe, if they are compensated by local variations in gravitational field energy, then it seems that macroscopic event horizons could actually be present even if matter was created out of nothing, as long as opposite action particles are not themselves inhomogeneously distributed in space. In fact, in the absence of an independent constraint that would impose a high degree of homogeneity on the matter distribution that emerged from the past singularity it could not even be required that all matter be created out of nothing during the Big Bang, because some of this matter could have survived a quantum bounce if it is contained in macroscopic black holes with opposite energy signs (because under such conditions it would not be submitted to opposite action pair annihilation). Therefore, even if we assume that the total energy of matter must have been nearly zero initially this does not provide sufficiently strong a constraint to require that the initial matter distribution be as highly homogeneous as it is observed to be.

It is certainly appropriate, however, to assume that in the more general context of a quantum theory of gravitation, the initial singularity merely constitutes a state of maximum positive and negative energy densities which would not be associated with an end of time, but which would rather give rise to a quantum bounce following which (in the past direction of time) space would stop contracting and begin expanding just as it did following the same event in the future direction of time. As I just mentioned though,

if time can be extended past the initial singularity in such a way, then it would seem possible for matter to have already been present before the time at which the singularity is formed and in such a case it would not necessarily need to be created out of nothing as pairs of opposite action particles, even in a zero energy universe. This is a simple consequence of the fact that an inhomogeneous distribution of opposite action matter particles would not be submitted to the kind of annihilation to nothing that would otherwise be allowed to occur in the instants preceding the formation of the past singularity (in the future direction of time), so that a large portion of the particles already present would be allowed to survive the quantum bounce.

Indeed, processes of opposite action pair annihilation to nothing can only occur when the particles involved are allowed to approach one another to within a quantum gravitational unit of distance, as I explained in section 1.9, and therefore they would not be allowed to take place if opposite action particles were originally concentrated in different portions of space and were being kept isolated from one another by macroscopic events horizons associated with black holes of opposite energy signs. The question, therefore, remains. Why was the distribution of matter energy so uniform in the very first instants of the Big Bang? I believe that one can only begin to understand the cause of the homogeneity of the matter distribution that emerged out of the past singularity when one acknowledges that what is significant about this initial state is the fact that it is characterized by a maximum matter density, because this is the only observable aspect of our universe which is correlated with the state of minimum gravitational entropy. What is significant, then, is that the cosmic horizon begins to grow at the exact moment when the density of matter is maximum. But why should causality have anything to do with the magnitude of the average densities of positive and negative energy matter?

I must admit that I always had difficulty accepting the very validity of the notion that the universe could have come into existence as a set of disconnected entities not causally related to one another due to the presence of multiple causal or cosmic horizons in the primordial state. The conclusion that the limited velocity of causal signals would forbid interactions between sufficiently distant regions of the universe, however, appeared unavoidable. But how could such an assortment of disconnected parts as is usually assumed to exist at the Big Bang be considered to form a single universe if its elements are not even related to one another in any way? How could they even have been allowed to come into contact with each other in a well-defined

manner later on if they weren't part of the same causally interrelated ensemble initially? This situation would be particularly puzzling in the context where we would consider that the Big Bang really constitutes the beginning of time, as there would then be no prior state at which causal relationships could have been established between the initially disconnected regions. Here again I just couldn't understand the appropriateness of a picture that most people accepted as valid without a second thought. But this led me to develop a better understanding of the conditions imposed by the principle of local causality on the initial Big Bang state that turned out to be crucial for explaining the high degree of homogeneity of the primordial matter distribution that is responsible for the existence of the thermodynamic arrow of time.

First of all, I think that it is important to mention that the notion that the size of the cosmic horizon increases with time, as the universe itself expands, contains an implicit assumption that is not always recognized for what it is. Indeed, when one considers that the horizon encompasses an increasingly larger portion of space in the future, one is actually presuming the validity of the classical principle of causality, that is, of the idea that causes always precede their effects. But it is actually always *past* causes that produce *future* effects. It is never assumed that a future cause could produce an effect in the past. This is usually appropriate, as we experience time in a unidirectional way as a consequence of the fact that the thermodynamic arrow of time always operates from past to future and never in the opposite direction. But when we are considering that no signal was allowed to propagate farther than the distance reached by the cosmic horizon at any given time after the Big Bang, we are implicitly assuming that it is only the past that can influence the future and that effects propagate in the future direction of time from causes which originate in the initial singularity. In other words, we are assuming the existence of a preferred direction in time (the future) and a preferred instant (the time at which the past singularity is formed) at which causes begin to propagate. But it must be clear that this is an assumption and that there is no a priori reason not to assume that classical causality instead operates toward the past from the instant at which a hypothetical future Big Crunch singularity would be formed, in which case the size of the horizon would already encompass all of space or at least a very large portion of it at the Big Bang.

The truth is that causality could begin to operate at any given instant of time, even the present time, and propagate from there in any given direction,

even the past. Therefore, if what we are seeking to explain is the existence of a preferred direction in time, then we cannot simply assume the validity of the classical concept of a horizon expanding from the Big Bang in the future direction of time. We cannot claim that there is a problem with the homogeneity of the large scale matter distribution if this problem arises as a consequence of assumptions concerning the size of the horizon which are only meaningful in the context where there is a direction to causal signals which originates from this very same homogeneity. What we must provide is a consistent justification for the very validity of this particular choice of a horizon concept. We must explain why this particular state in the past was configured in such a way that it allowed classical (unidirectional) causality to be a meaningful concept that came into effect at the exact moment when matter emerged from the past singularity.

But even apart from those considerations, the cosmic horizon concept as it is currently understood is somewhat problematic in that, quite ironically, it does not provide any specific requirement for the existence of causal relationships among the various elements of the universe. Despite those difficulties I came to recognize the validity of the limitations imposed by the existence of cosmic horizons. I believe that what allows this concept to be acceptably formulated is simply the fact that, ultimately, as we consider increasingly earlier times, the size of the causal horizon would actually reach the limit imposed by quantum theory on the definiteness of any measure of spatial distance or area. When the size of the cosmic horizon reaches the limit in the past at which the non-locality that is intrinsic to quantum phenomena becomes prevalent, it is certainly appropriate to no longer require the limited velocity of causal signals to forbid the existence of physical relationships between the particles present within the size of the horizon at that moment, as everything within that horizon must be assumed to be connected to everything else (inside the same horizon) as a consequence of quantum non-locality and entanglement.

In such a context it would be sensible to assume that there may after all exist relations of causality between all physical elements of the universe which were in contact with one another to within an elementary quantum gravitational unit of distance at the Planck time, when the cosmic horizon was small enough that quantum indeterminacy could not be ignored. In fact, from a quantum gravitational viewpoint it may be preferable to simply recognize that there is nothing smaller than the elementary units of space associated with this particular scale. But given that causality is a feature of

the classical spacetime structure, this means that there would be no sense in imposing limitations on signal propagation below that scale. Therefore, when the size of the cosmic horizon reaches the natural limit imposed by quantum gravitation, a Planck time after the initial singularity, if the most elementary particles are allowed to be in contact with one another, then no smaller components would remain causally unrelated, which is probably sufficient a condition to impose regarding the necessity for the universe to form a globally consistent whole with all of its elementary particles having been in direct contact with another particle at least once, before becoming separated by large spatial distances.

Now, this simple formulation of the requirement which I believe allows the universe to exist as the ensemble of all those things which are physically related to one another and to nothing else may appear benign, even if adequate, but in fact it can be attributed the most amazing consequences in the context where it is recognized that negative energy matter must be assumed to exist in our universe. Thus, I would like to suggest that all the elementary particles originally present in the universe at the Big Bang and from which evolved the current matter distribution be required to have been in contact with at least one other particle at the Planck time. More specifically, I propose that the following condition must necessarily apply.

> **Global entanglement constraint**: There must exist an event at one particular moment of cosmic time when all the elementary particles which are then present in the universe, regardless of their energy sign, were in direct contact with at least one neighboring elementary particle of either positive or negative energy sign in a state of maximum matter density.

If this condition is fulfilled then any particle that is present in the universe would have once been in contact with a particle that was in contact with another particle and so on, which means that at no time could a physical element of the universe exist that would be causally unrelated to the other elements which are considered to be part of the same ensemble, even if the particles which were initially present in the maximum density state later become separated by spacelike intervals and are no longer in contact with one another. If this requirement was not fulfilled there would be no reason to expect that when the cosmic horizon grows in the future, particles which were causally unrelated initially could begin to influence one another through

long range interactions, because those particles would not even be elements of the same universe.

Of course the existence of such a smallest, physically significant cosmic horizon does not mean that the limits imposed by the size of cosmic horizons on the propagation of causal signals no longer apply, but merely that they need not apply at times earlier than the Planck time. It must be clear that there would be no sense in speaking about the ultimate horizon as being that which would be associated with the epoch at which the whole universe would be contained within a single Planck surface, because once there is a matter particle in every elementary unit of space and the average density of positive and negative matter or vacuum energies is maximum, no further contraction is possible, as all tentative quantum theories of gravitation appear to confirm. What this means is that it wouldn't even make sense to impose a condition of causal contact on a state that would be reached at an even earlier time. Thus, even if the constraint of global entanglement concerns the state of the universe at the Planck time it would be incorrect to assume that only the detailed knowledge of a fully developed quantum theory of gravitation would allow us to say anything meaningful regarding the state of the universe at such an early time. But we still need to explain why it is that the matter distribution was almost perfectly smooth, even on a scale larger than the size of the horizon at the time when the density was maximum, as required if the growth of this cosmic horizon, as a unidirectional phenomenon, is to actually begin at that particular instant of time. This is a particularly difficult question given that large scale homogeneity is precisely what would appear to be forbidden by the existence of such a horizon.

The implications of the global entanglement constraint only emerge in the context where it is recognized that event horizons (such as those associated with black holes) can under certain conditions constitute potential barriers which are impossible to overcome. It must be clear, first of all, that even though certain positive energy particles could be prevented from coming into contact with other positive energy particles in the initial state of maximum matter density as a consequence of being contained within the macroscopic event horizon of a positive energy black hole, if only positive energy matter existed this would not allow to justify imposing a limit on the amplitude of primordial density fluctuations, because in such a case, regardless of the presence of macroscopic event horizons, all matter particles would eventually end up being in contact with their neighbors as the contraction of space that takes place backward in time toward the initial Big Bang state would lead

to the merger of all the event horizons which were originally present and their spacetime singularities, as in a generic Big Crunch process. Under such conditions all particles which may now be isolated by the presence of event horizons would nevertheless merge into one initial state of maximum matter density where every particle would occupy an elementary unit of space and be in contact with all the surrounding particles in this initial singularity. Thus, if only positive energy matter was present in our universe it would seem that the global entanglement constraint could be satisfied in the initial state without gravitational entropy being minimal, because even if strong local gravitational fields associated with the presence of event horizons could exist in the instants immediately preceding the formation of the singularity (in the past direction of time), all elementary particles would nevertheless be allowed to be in contact with their neighbors in the maximum density state, because those are attractive gravitational fields.

When negative energy matter is present, however, things become more complicated. Indeed, if the constraint of global entanglement imposes contact between neighboring elementary particles at the Planck time, regardless of their energy sign, then given that gravitational repulsion, unlike gravitational attraction, may forbid local contacts by giving rise to insurmountable potential barriers for particles located within black hole singularities of opposite energy signs, it follows that event horizons can be expected to be absent initially on all but the smallest scale, even if macroscopic black holes are allowed to form at later times. If this was not the case then certain particles could exist in our universe that would not be causally related to the rest of it, which I believe would involve a contradiction. In the absence of a condition of global entanglement the most likely initial state, from a purely statistical viewpoint, would be one for which all the matter in the universe would be concentrated in the smallest possible number of opposite energy black holes with arbitrarily large masses which would already be in a state of maximum gravitational entropy. But this was not allowed to constitute our boundary conditions at the Big Bang simply because under such conditions the singularities at the center of the objects could never come into contact with one another in the maximum density state, while this is required by the global entanglement constraint.

In the presence of negative energy matter global entanglement actually constitutes a very constraining requirement, because when the average densities of positive and negative matter energy are so large that they reach the theoretical limit imposed by the magnitude of the natural vacuum-stress-

energy tensors, the slightest local overabundance or rarity of positive matter energy in comparison with negative matter energy would give rise to the presence of event horizons which would forbid the condition from applying. Therefore, such density fluctuations must be nearly completely absent in the first instants of the Big Bang and can only develop gradually at later times, in an initially smooth and homogeneous distribution of positive and negative matter energy. The mass of any black hole that is now present in the universe should therefore diminish continuously in the past direction of time as we approach the initial singularity, so as to allow the condition imposed on the initial state to be satisfied, despite the fact that it is actually the past condition that gives rise to the future configuration in the context where the condition that applies on the initial Big Bang state is indeed one of minimum gravitational entropy from which the classical (unidirectional) principle of causality itself can be expected to emerge.

What was so puzzling about the previously unexplained fact that an ever smaller number of microscopic configurations seems to be available for matter evolving in the past direction of time under the influence of the gravitational interaction was that no such a decrease in the number of allowed microscopic configurations is observed in the future direction of time. As a consequence of this limitation, predictions of a statistical nature, such as those made using quantum theory, are always valid only for evolution toward the future while evolution toward the past cannot in general be accurately predicted (the probability of prior events cannot be determined from that of subsequent events, while the probability of future events can usually be determined from that of past events), which is annoying given that the equations of the theory are symmetric under a reversal of the direction of time. But this is not a consequence of the fact that information concerning the state in which a system will evolve is only available for the past and not the future, because it is possible to recognize retrospectively the absence of statistically significant constraints that would apply to future evolution by considering the future of an initial state at a time when this future is now itself in the past. This is in contrast with the evolution that can be observed to take place at the same time toward the past and which reveals that systems can only come to occupy a subset of their theoretically allowed microscopic configurations whose only distinctive property is its lower entropy.

What remained unexplainable, therefore, is the fact that an ensemble of systems started in the same macroscopic state evolve to occupy all available microscopic states in the future, while a similar ensemble started in the

same macroscopic state usually evolve only to past states characterized by a lower entropy and in particular a lower gravitational entropy. But I have now explained that this diminution in the number of available microscopic states toward the past originates from the necessity that all the elementary particles present in the maximum density state at the Big Bang come into contact with their neighbors of any energy sign in order that there exist causal relationships between all independently evolving components of the universe. The unnatural evolution that takes place in the past is the direct consequence of the limitation imposed on the initial state by the condition of global entanglement in the presence of negative energy matter and it would not merely characterize a small portion of all possible universes, but really all universes governed by the known fundamental principles of physics. Remarkably enough, this condition allows to explain why it is that only the gravitational component of entropy was not maximum at the Big Bang, while the entropy of matter and radiation was allowed to be arbitrarily large, which is certainly appropriate given that the universe was then already in a state of thermal equilibrium. The constraint of global entanglement only limits the magnitude of entropy attributable to the gravitational field and this is exactly what we need.

It must be clear that the fact that a perfectly uniform distribution of negative energy matter exerts no gravitational influence on positive energy matter does not allow one to assume that it is not necessary to take into account the presence of negative energy matter in trying to identify the origin of the constraints that give rise to the homogeneous initial distribution of positive matter energy, because it is precisely the magnitude of local inhomogeneities in the distribution of positive and negative matter energy which needs to be constrained and negative energy matter inhomogeneities do have an effect on positive energy matter, particularly when the average matter density is maximum. In fact, negative energy matter always exerts an influence on positive energy matter under conditions of maximum average matter density, because locally elementary black holes are necessarily present (as I have mentioned while discussing the problem of black hole entropy in section 2.11) and the energy distribution is never perfectly smooth and homogeneous, especially in the context where it is understood that two particles with maximum opposite energies cannot be located in the same elementary unit of space, due to the insurmountable gravitational repulsion they would exert on one another.

The constraint of global entanglement, therefore, merely imposes that the positive and negative matter energy present in the initial maximum den-

sity state be as homogeneously distributed as necessary for the absence of *macroscopic* event horizons associated with black hole masses larger than the Planck mass, because it is only under such conditions that the most elementary particles of matter (with the highest possible positive and negative energies), submitted to the gravitational fields of the most elementary black holes (with the smallest possible surface areas), can be in direct contact with one another regardless of their energy sign and thus be part of the same universe. This conclusion remains valid even in the context where it must be considered that there is no direct interaction between positive and negative energy particles under normal conditions, because on a sufficiently short scale the indirect gravitational interaction between opposite energy particles is strong enough to allow them to exert an influence on one another.

While the event horizon of a macroscopic negative energy black hole may prevent local contact between the particles that reached its singularity and neighboring positive energy particles which cannot cross this event horizon, an elementary black hole, by virtue of its minimum size, would merely constitute the surface of the one and only elementary particle whose motion it constrains, which means that this particle would be allowed to come into contact with particles which are under the influence of the gravitational fields of other elementary black holes in the state of maximum matter density, regardless of their energy signs. This is what justifies assuming that the condition of global entanglement only imposes an absence of *macroscopic* event horizons. If all the matter in the universe was initially concentrated in two macroscopic black holes of opposite energy signs, the particles contained in the singularity of one of the object would remain isolated from those contained in the singularity of the other black hole, even if the event horizons of the two objects were in contact with one another and this is what must be considered forbidden by the constraint of global entanglement. The very meaningfulness of this condition is in fact dependent on the hypothesis that there exists a minimum physically significant spatial scale below which no causal signal needs to have propagated and which corresponds precisely to the size of an elementary black hole which is associated with the state of at most one particle of maximum positive or negative energy.

What's interesting is that contrarily to the situation we would have if inflation was assumed to be responsible for the smoothness of the initial matter distribution, it is now possible to explain why it is that the constraint that gives rise to a homogeneous initial state is necessarily effective in only one direction of time. Thus, gravitational entropy can be expected to decrease

continuously in the past direction of time from its current intermediary value, even if this would appear to be a very unlikely evolution for the universe to go through from a statistical viewpoint, because if the present inhomogeneity is not reduced then the smooth merger of the positive and negative energy matter distributions that is required for the global entanglement of all particles to take place would not happen. This reduction of gravitational entropy can now be understood to occur regardless of whether space is expanding or contracting, as long as we are in effect approaching the instant at which is formed the unique singularity on which the condition of global entanglement is to be imposed.

It is, therefore, simply the fact that the condition that applies to the initial singularity is precisely one of minimum gravitational entropy from which can emerge a phenomenon of classical (unidirectional) causality that operates toward the future from that particular instant of time, that requires the evolution that takes place at all later times to be such that it allows an initial state obeying this condition to be reached in the past direction of time, because under such circumstances it is in effect past conditions that determine future evolution. If quantum theory only works for predicting future events it is because all possibilities are indeed allowed for evolution toward the future, while only a limited subset of potentialities can be actualized in the past as a consequence of the constraint that is continuously being exerted on past evolution by the requirement of global entanglement, which imposes a low gravitational entropy on the state which existed in the past when the density of matter was maximum. It is quite remarkable that this apparent backward teleology can be shown to arise from the existence of an inescapable constraint that applies on one particular state only, but even more surprising is the fact that this can actually be achieved through the application of fully time-symmetric physical laws which gave no hint of having the potential to produce such a manifestly irreversible evolution.

It is important to emphasize that in the context of this explanation of temporal irreversibility all physical systems, regardless of how isolated they may have become at the present time, must evolve with continuously decreasing gravitational entropy in the same past direction of time, because they are all submitted to the same unavoidable constraint applying to the same unique state of maximum matter density in the past. This constraint, therefore, is stronger than any condition that would be imposed independently on the present state of one or another system in order to favor an evolution to lower entropy states in a given arbitrarily chosen direction of time. Indeed, in the

context where all processes are fundamentally unpredictable, a constraint that would apply merely to the present state of a non-equilibrium system could not alone impose on this system that it evolves with decreasing gravitational entropy over a very long period of time in either the past or the future, regardless of how carefully the system is prepared. This is in contrast with the constraint imposed by the condition of global entanglement which by definition must necessarily and unavoidably apply to all physical systems which are part of the same universe and of no other and which exerts its influence incessantly in the same unique direction of time (toward the initial singularity) and in such a way gives rise to an asymmetry which is actually shared by all systems, including any branch systems which are no longer in contact with their environment. In the present context this temporal parallelism is a simple consequence of the fact that all physical systems in the universe are lead by a common condition which applies to the state they occupied when the cosmic horizon began to spread and which originates from the requirement that they actually be part of the same universe.

If the initial or final conditions applying on current states cannot alone explain the temporal parallelism of branch systems it is because even if this would be possible for the evolution that takes place in the future direction of time, the fact that for all practical purpose such isolated systems never evolve toward a state of higher gravitational entropy in the past direction of time, like they do in the future, but rather *always* evolve to even lower entropy states in the past, means that it is not the conditions applying on current states which alone determine their past evolution. It is precisely the fact that the requirement of global entanglement must, as a matter of consistency, apply to all particles in the universe that guarantees that all branch systems without any exception must obey the same constraint of decreasing gravitational entropy in the same direction of time toward the initial singularity. The parallel thermodynamic behavior of isolated branch systems can be expected to occur as a result of the fact that any system that is part of a given universe, regardless of how isolated it might have become, must have been entangled with the rest of the matter in this universe at the Big Bang in order that causal relationships be established between all components of the universe and this implies that even those portions of the universe which are now isolated must follow the same kind of gravitational entropy decreasing evolution that is necessary for achieving this global entanglement at some point in the past.

The parallelism of the asymmetry of thermodynamic evolution can only

be explained if there exists a constraint that requires the diminution of the gravitational entropy of all systems in the past direction of time independently from what their initial or final states are at the present time and the fact that such parallelism is actually observed under all circumstances clearly shows the validity of the arguments that allowed me to determine the nature of this constraint. If those considerations are appropriate, then it would mean that the assumption that the initial conditions at the Big Bang should always be fixed arbitrarily, which would appear to conflict with the assumption that thermodynamics is fundamental and irreducible, is not really incompatible with the notion that there exists a constraint applying on those initial conditions that gives rise to irreversible thermodynamic evolution as a derived property.

What is important to understand is that a maximum density state must necessarily occur at one time or another for the global entanglement of all elementary particles to be satisfied and given that such a state would not likely be characterized by an absence of macroscopic event horizons unless it constitutes the mandatory unique event at which global entanglement is enforced on the universe, then one must conclude that our Big Bang really is this unique event. In such a context the presence of an initial singularity would no longer be a mere fortuitous consequence of the fact that space is expanding, but would be an essential requirement for the existence of any universe obeying the known principles of physics. I believe that it is the widespread ignorance of this fact that explains that it took so much time for all the consequences of the presence of a Big Bang to be properly understood and appreciated. To the usual three pieces of evidence in favor of the Big Bang which are the observation that space is expanding, the accuracy of the prediction of light element abundances, and the detection of the cosmic microwave background, I would therefore suggest that one adds the theoretical argument concerning the very necessity of a maximum density state, which is made conspicuous by the undeniable character of our experience of a thermodynamic arrow of time.

Once we recognize that there actually exists an independent requirement for an initial state of maximum matter density, then any attempt at explaining its apparent unlikeliness by postulating that it emerged from a fluctuation that occurred in a maximally extended empty space can no longer be considered satisfactory. Even if such a scenario was found to work as advertised it would remain inconsistent from the viewpoint of the principle of local causality, unless we assume that this extended empty space itself emerged out of

another creation event that would have taken place at an earlier time when the matter density was actually maximum and global entanglement was allowed to take place. Thus, comments to the effect that it would become impossible to explain the existence of an arrow of time if there only existed one single universe in all of space and one single Big Bang in all of time appear to be misguided, because in fact it seems that the truth is to be found in the exact opposite statement. It is as a result of having tried very hard to understand why it is that there should in effect be a unique initial state of maximum matter density for the universe, by first acknowledging that this is a perfectly legitimate hypothesis, that I was allowed to achieve progress in identifying the cause of the homogeneity of the initial distribution of matter and energy that gave rise to temporal irreversibility.

If gravitational entropy does indeed rise in only one particular direction of time it is because only evolution away from the initial singularity, either in the future or in the past, can be expected to be left unconstrained by the condition of global entanglement, which actually gives rise to a well-defined thermodynamic arrow of time independent of whether space is expanding or contracting. It is, therefore, possible to understand why it is that classical causality operates from past to future in the portion of history that follows the initial singularity and also why it is that the cosmic horizon only begins to spread outward at the Big Bang. It is the fact that the condition of global entanglement would only be required to apply once, even if the universe was to return to a state of maximum matter density at some point in the future, that explains that the evolution that takes place from the moment at which this condition is enforced is not symmetric in time. Thus, it is incorrect to argue, as certain authors do, that in order not to assume the very outcome we are seeking to derive (the temporal irreversibility) it is required that any condition that applies to some initial state should also apply to a final state. Once it is understood that there need only be one state of high matter density and low gravitational entropy in any given universe, then the kind of evolution which can be expected to take place in the direction of time toward that unique state, either in the past or in the future direction of time, would necessarily be different from that occurring in the opposite direction and this allows to explain time asymmetry without assuming it in the first place. What I'm assuming here is not the asymmetry itself, but merely the uniqueness of the state which allows it to arise. I'm not picking up a unique direction of time, I'm merely identifying the necessary conditions that must apply to the distribution of matter energy at one single moment of time and

it just happens that those conditions are so unlikely to ever be satisfied again by chance alone that any later or earlier evolution can be expected to take place irreversibly.

Now, if *bidirectional* time does extend past the 'initial' singularity following a quantum bounce, we can expect space to be expanding and the density of matter to be decreasing immediately after the event (in the past direction of time), while the inhomogeneity of the matter distribution would still need to be minimum if there is to be any continuity in the evolution of the microscopic state of matter and its gravitational field as we pass the point of maximum positive and negative energy densities. But this means that, even for this portion of history, the thermodynamic arrow of time would initially have the same direction as the cosmological arrow of time associated with expansion and would actually be opposite that we observe on our side in time of the initial singularity. As a result, the area of black hole event horizons and the associated gravitational entropy would be growing toward the past (which any observer then present would consider to be her future), which means that in the future direction of time the same objects would evolve as white holes emerging from generic (high entropy) past singularities. Thus, it would be inappropriate to simply propose that it is because a condition of low gravitational entropy applies to *all* past singularities that the energy of the matter that emerged from the Big Bang was so uniformly distributed. Anyhow, it must be clear that if the thermodynamic arrow of time is indeed reversed as soon as the instant of the initial singularity is reached, then whatever occurred during the portion of history that preceded the Big Bang would remain unknowable to observers in the current portion of history. This would be true for the exact same reason that events located in our future cannot be known in advance, which is attributable to the fact that classical (unidirectional) causality and the formation of mutually consistent records of events only take place in the direction of time relative to which entropy is rising.

Still regarding the possibility for bidirectional time to extend past the initial singularity, I believe that it would be inappropriate to assume that if this hypothesis is valid it would become impossible to explain the low gravitational entropy of the Big Bang state by imposing a condition on the initial singularity. It was argued in effect that if there was a history prior to the Big Bang, the final singularity which would be produced in the future direction of time (which would constitute our initial singularity) would likely be in a high gravitational entropy state (as any future state reached after a

long period of random evolution), which would require the state following it (our initial state) to have a similar configuration. But in fact, it is exactly the opposite which is true and the state preceding the initial singularity must actually be very homogeneous, because the constraint of global entanglement applies to the singularity itself, while it is the evolution away from it in *any* direction of time which is unconstrained. Continuity merely imposes that the configuration be similar on both sides of the initial singularity, but it does not allow one to determine what this configuration actually is. It is in effect only in the absence of an appropriate constraint to be imposed on the initial singularity that gravitational entropy would have to be maximum in both the immediate past and the immediate future of the initial singularity and indeed at all times. Not recognizing this would again amount to favor one particular direction of time (that relative to which entropy would be assumed to grow before the initial singularity) without justification, instead of explaining why such a preferred direction naturally emerges, as I have done.

What's interesting is that when time is actually unfolding in such a way past the Big Bang singularity, then the universe is allowed to be completely symmetric with respect to past and future, because in the presence of negative energy matter not only is there symmetry under a reversal of the fundamental direction of propagation in time of elementary particles associated with the distinction between matter and antimatter (as I proposed in section 2.10), but on a global scale there is also invariance under a reversal of the thermodynamic arrow of time. If, in addition, the matter-antimatter asymmetry can be assumed to be reversed for positive or negative action matter (independently) in that portion of history unfolding past the Big Bang singularity (as a result of the fact that the C symmetry actually involves a reversal of time from a bidirectional viewpoint), then this extension of time would also allow the universe to regain the symmetry that would otherwise be lost as a consequence of the fact (discussed in section 2.10) that there must be a larger number of positive energy (not positive action) particles propagating in any direction of time in the known portion of history. Thus, it is perhaps not just possible, but actually compulsory to assume that there is in effect a history not so unlike our own that unfolds in the past direction of time before the instant of the initial singularity. In any case, if a quantum bounce does occur for the entire universe when the state of maximum matter density is reached in the past, this process would not violate the rule of entropy increase as would an ordinary white hole, because what we are considering here is not the time-reverse of a generic Big Crunch gravitational collapse and in

the present context the thermodynamic arrow of time associated with the variation of gravitational entropy would itself reverse at the exact moment when the universe begins expanding.

The picture that develops is, therefore, that of a universe for which gravitational entropy is growing continuously in both the future and the past of a state of maximum matter density in which the distribution of matter energy was almost perfectly uniform. This evolution is taking place from both the viewpoint of positive energy observers and that of negative energy observers and can be expected to continue regardless of whether space is expanding ever more rapidly or eventually begins to recontract. In the context where gravitational entropy can be expected to grow as a consequence of the dissociation of the positive and negative energy matter distributions it follows that if there is an infinite amount of matter in the universe there may never arise a state of maximum stability, equivalent to thermal equilibrium, where gravitational entropy would become arbitrarily large and would no longer rise. Under such conditions it cannot be expected that the universe will ever evolve back to a state similar in every respects to the state in which it was at the Big Bang, because the probability that such a universal Poincaré return would occur is not merely low, it is decreasing all the time. We may, therefore, be justified in describing the evolution that takes place on a cosmic scale in both the very far future and the very far past as truly irreversible. The ancient view of a universe reaching its heat death and remaining in this sterile randomly fluctuating state forever may well be incompatible with the most basic theoretical constraints governing its birth process and later evolution which rather bespeak of its potential for eternal vitality.

Returning to the problem of matter creation, it is now possible to understand that in the context where every matter particle in the universe must be causally related to the rest of it, all the matter present in the universe need to be produced out of nothing as opposite action pairs during the first instants of the Big Bang, even if time does extend past the initial singularity. This is a consequence of the fact that when positive and negative action particles are as homogeneously distributed as required by the constraint of global entanglement in the state of maximum positive and negative energy densities that immediately precede the quantum bounce (in the future direction of time), all positive action particles are allowed to annihilate to nothing with a nearby negative action particle and therefore no matter remains that would have been produced before the past singularity, which means that all the

matter that emerges from the singularity must have been created out of nothing during the Big Bang. What's more, even if the constraint of global entanglement was required to apply to any delayed matter creation event (for the same reason that justifies that it be imposed to the Big Bang), matter creation out of nothing would not be allowed to occur again once the expansion rate has been slowed down by the early process of matter creation itself, as I explained in section 3.5, and therefore all matter, regardless of how homogeneously distributed it might be, must definitely be the product of one single, genuine Big Bang.

But, even though the *energy* of the matter created out of nothing in the first instants of the Big Bang must be as homogeneously distributed as the matter particles themselves, the constraint of global entanglement does not impose a perfect homogeneity, while the principle of conservation of energy and the requirement of relational definition of physical attributes only require the universe to be flat and the energy of matter and vacuum to be null on the scale of the universe as a whole and does not forbid the density of negative energy matter to differ from that of positive energy matter locally (even though the weak anthropic principle would actually prevent those densities from differing significantly on the global scale). As a result, despite the overall uniformity of the matter distribution, there can exist small differences between the magnitudes of the densities of positive and negative matter energy on all scales in the initial state that emerged from the Big Bang singularity and it is those fluctuations that would give rise to present-day structures.

Now, in the context where both positive and negative matter energies were in effect uniformly distributed in the first instants of the known Big Bang, as a consequence of the requirement of global entanglement, if the rate of expansion must be critical in order that the universe have a null energy, as I proposed in section 3.5, then it becomes possible to conclude that the universe must expand isotropically to a very high degree of precision, even in the absence of an initial phase of inflationary expansion, because under such conditions the expansion rate is roughly the same everywhere. But if the expansion actually is isotropic around every point, as a consequence of the requirement that the expansion rate be fixed by the matter density, then it follows that the matter distribution must remain homogeneous on the largest scale as expansion proceeds. As a consequence, it is plausible to assume that the Big Bang happened everywhere at once at nearly the same instant of cosmic time, which once again allows to exclude the possibility that there

might have occurred delayed creation events.

The uniformity of the expansion rate also allows one to deduce that the temperature of the cosmic microwave background should be homogeneous even on a scale larger than the size of the cosmic horizon at the epoch of recombination, because the absence of macroscopic event horizons is required on all scales and this imposes very stringent conditions on the fluctuations of matter energy density that could be observed, even on the largest scale. In fact, the condition of global entanglement can be expected to exert an even greater constraint on the magnitude of fluctuations in the density of matter occurring on a larger scale, given that an overdensity of lower magnitude would be required to produce a macroscopic event horizon on such a scale, as witness the fact that larger black holes have lower mass densities. This means that no smoothing process is required to make the temperature of the cosmic microwave background uniform, because the distribution of matter energy and the expansion rate were mostly uniform on all scales right from the beginning, even if the size of the cosmic horizon decreases more rapidly than the scale factor as we approach the initial Big Bang singularity in the past direction of time, so that regions which are now in contact must have been causally disjoint at the epoch of recombination.

When one properly recognizes the limitations imposed by the global entanglement constraint on the initial state at the Big Bang, the horizon problem simply no longer exists and no independent assumption is required to confirm the relevance of the cosmological principle for a description of the early universe. There is no longer any mystery associated with the fact that only one parameter (the scale factor) is required to describe the state of the universe at all but the most recent epoch. In fact, it would now appear that the cosmological principle must be obeyed as accurately as we are considering increasingly larger regions of space, at times increasingly closer to the initial singularity. Despite the enormous densities and the extreme temperature that characterizes the Big Bang it would therefore be possible to determine the general properties of the initial state with much greater precision than is usually assumed, despite an absence of knowledge of the exact unified theory that would apply under such conditions. To be sure, the usual assumption that in order to obtain a homogeneous matter distribution on the cosmic scale it is necessary for the entire observable universe to have been contained within the cosmic horizon at some point in the past (which would be impossible without inflation) can now be recognized as inappropriate and unnecessary.

In the context where it is indeed the magnitude of local fluctuations in the primordial distribution of matter energy which is restricted by the global entanglement constraint, it would also follow that arguments to the effect that topological defects should have been abundantly produced at the Big Bang may no longer be as significant as they used to be. Of course, even from a conventional viewpoint one must be careful when considering the prediction that there should have formed in the universe a large number of magnetic monopoles or cosmic strings, because the validity of the grand unified theories on which those deductions are based hasn't yet been experimentally confirmed. However, some of those predictions appear to be largely independent from the details of the theories from which they are derived and therefore cannot be ignored. What I have realized is that the relatively low abundance of topological defects may simply be a consequence of the fact that they are very high energy objects, similar in certain respects to naked singularities of the future kind. The presence in the initial state of compact objects that would concentrate such large amounts of positive or negative energy in such small volumes of space may simply be incompatible with the requirement of smoothness of the primordial distribution of matter energy and the absence of event horizons that is imposed by the constraint of global entanglement. Indeed, magnetic monopoles are sometimes described as magnetically charged black holes and if this characterization is appropriate it would certainly follow that no such an object could be present in the maximum density state at which global entanglement is effected.

It is true, however, that it is only the presence of such objects in the very first instants of the Big Bang that can be ruled out on the basis of the requirement of global entanglement. If topological defects were expected to arise only at later times, then this condition alone may not forbid their existence. But in such a case the rarity of topological defects may simply be a consequence of the fact that the amplitude of fluctuations in the energy density of matter is too small initially to allow the production of highly dense topological defects at later times. I believe that this limitation, which simply does not exist from a traditional viewpoint, actually provides the strongest and most unavoidable limitation regarding the presence of topological defects. In any case, the fact that the vacuum itself is a much different phenomenon in the context where its natural, average energy density is actually null certainly contributes to explain why it is that the traditional expectations regarding the cosmological consequences of symmetry-breaking phase transitions are not reflected in what we already have of experimental evidence. Those ex-

planations may not be as satisfactory as the solutions I have provided to other aspects of the inflation problem, but given that according to the most knowledgeable experts there is only a very small chance that the conditions could be met that would allow inflation to occur and to last for a sufficiently long time that it could actually reduce the density of topological defects to acceptable levels, then we may have no choice but to recognize that the constraint discussed above provides a more solid foundation for explaining the rarity of those theoretical objects. In any case, the fact that the physics of topological defects is still relatively uncertain means that the tentative explanation provided here cannot be rejected, even if at this point it is not itself entirely conclusive.

Now, it is sometimes argued that the distribution of matter energy was so uniform at the time when the cosmic microwave background was released that what remains unexplainable is really that the temperature was not perfectly smooth and free of any fluctuations initially. But I believe that this smoothness problem is a mere consequence of the fact that we do not properly understand what gives rise to the high level of uniformity of the initial distribution of matter energy. It is only when this smoothness is assumed to be perfect by default that we must invoke a cause in order to explain the fact that there actually existed fluctuations in the density of matter energy in the past on distances larger than the size of the horizon. Given that in the context of the above discussed solution to the horizon problem it is merely the upper bound of fluctuations in matter energy density which is constrained, then it is to be expected that certain local variations in matter energy density would necessarily be present, as the absence of macroscopic event horizons can be satisfied even when certain fluctuations are present and therefore if the initial state is still chosen randomly, as it should, it would likely not be perfectly smooth. We do not need causal influences and the propagation of local perturbations to give rise to the fluctuations observed on the largest scales in the cosmic microwave background. The cause of the correlations between the variations in the density of matter energy occurring in regions of the universe larger than the horizon is the constraint of global entanglement itself, which also requires a certain local smoothness in the inhomogeneities, thereby giving rise to the presence of structures above the horizon size, which need not have involved the propagation of causal influences as the allowed inhomogeneities were already present initially. The fact that there is smoothness even in the fluctuations is explained by the nature of the constraint responsible for the overall homogeneity. It is, therefore, the

very condition of global entanglement which is responsible for the fact that there does exist regularities in the fluctuations which are otherwise allowed to be present.

The usually favored approach to the problem of the origin of primordial inhomogeneities in the distribution of matter energy, which involves assuming that they arise as a consequence of irreducible quantum fluctuations initially present in the distribution of vacuum energy, only makes sense in the context where inflation is assumed to generate an otherwise perfectly homogeneous 'initial' state out of a much smaller volume of space. From my viewpoint what must be explained is not the presence of inhomogeneities, but the overall uniformity which, in the absence of a specific constraint, should not be observed. The natural configuration for the initial state is not one of perfect smoothness and there is no need to invoke a particular effect to generate the observed fluctuations, which are allowed to be present as long as they do not violate the condition of global entanglement. What is truly remarkable is that the spectrum of matter energy density fluctuations which is actually deduced from observations of cosmic microwave background temperature fluctuations is a scale-independent spectrum of the Harrison-Zel'dovich type (for fluctuations larger than the scale of the horizon at the time of recombination) while this is the only spectrum which according to specialists does not give rise to the creation of a large number of primordial black holes on smaller scales, nor to large deviations from homogeneity on larger scales, and those are precisely the conditions which are required by the theoretical constraint of global entanglement.

It should be clear, anyhow, that negative energy matter does have an effect on the observed properties of cosmic microwave background temperature fluctuations. Of all the measurements concerning the spectrum of CMB fluctuations, the only ones which would remain mostly unaffected by the presence of negative energy matter are those which regard a determination of the angular scale of fluctuations from which are derived the average density of positive matter and vacuum energy, because the trajectories of positive energy photons are not affected by the presence of a homogeneous distribution of negative energy matter. This appears to be confirmed by the fact that estimates of the density of positive energy matter (both visible and dark) based on measurements of the spectrum of CMB temperature fluctuations produce a value largely equivalent to that which is derived using more direct methods. I must acknowledge, however, that discrepancies have emerged more recently [40, 41], which may be resolved once it is recognized that the pres-

ence of negative energy matter must affect the calculations from which one derives the size of the sound horizon that is observed in the visible portion of the cosmic microwave background, thereby allowing one to avoid having to assume that the universe cooled more rapidly due to an unexpectedly large early expansion rate, as would otherwise appear necessary.

But it can also be expected that the inhomogeneities which are present in the early distribution of negative energy matter would actually alter the amplitude and the distribution of CMB temperature fluctuations to a certain extent, because those inhomogeneities should be as developed as those which are present at the same epoch in the positive energy matter distribution. Given that what must happen in the presence of negative energy matter inhomogeneities is more temperature fluctuations, then it would appear necessary to revise the magnitude of density fluctuations attributable to positive energy matter downward. Yet, given that negative energy matter interacts with positive energy matter only through the very weak gravitational interaction, it is possible to conclude that the inhomogeneities which are present in the distribution of negative matter energy should not affect the spectrum of CMB temperature fluctuations as much as those which are present in the distribution of positive matter energy, which means that only relatively small corrections would need to be applied to earlier estimates of the value of cosmological parameters based on measurements of those temperature fluctuations.

It is certainly possible, therefore, that some characteristic features of CMB temperature fluctuations which remain unexplainable in the context of current cosmological models could actually be explained by taking into account the effects that would be attributable to the presence of inhomogeneities in the distribution of negative matter energy, either in the initial state, or in the space through which radiation propagates before reaching our detectors. I have already mentioned in section 3.4 the possibility that the observed alignment of CMB temperature fluctuations along a certain axis in space could actually be the consequence of the presence of a very large scale inhomogeneity in the primordial distribution of positive and negative matter energies, but other more subtle effects are certainly possible as well. It may in particular be the case that some unexplained anomalies in the amplitude of CMB fluctuations observed on the largest scales are attributable to the variation of the cosmological constant which is expected to take place according to the developments introduced in section 3.2. One very clear implication of a cosmological model based on the ideas developed in this report, however,

is that we will never observe the expected gravitational wave signal which according to traditional models should show up in the polarization of CMB radiation in the context where the smoothness of the primordial distribution of matter energy is produced by inflation, starting from a highly inhomogeneous initial state. This must now be considered unavoidable, because even if there occurred an early phase of accelerated expansion during the first instants of the Big Bang, the initial state preceding it must have been almost perfectly uniform already and no gravitational waves would be produced as a result of its stretching.

3.10   A criticism of inflation theory

Now that I have provided alternative solutions to all aspects of the inflation problem I would like to offer a constructive criticism of inflation theory itself and explain why it may no longer constitute an appropriate response to the most enduring difficulties facing theoretical cosmology. It must be clear, however, that I do not claim to have proven that inflation theory is wrong or that the phenomenon it describes did not occur. Indeed, what I showed is simply that inflation is no longer *necessary* to solve the flatness problem and the related problem of matter creation and that an alternative solution to the horizon problem and the related problem of smoothness can be formulated that may also go some way in solving the problem of topological defects. But this does not mean that the hypothesis that there occurred an early phase of accelerated expansion is not valid and that we should no longer expect something like inflation to have happened, only that the existence of such a phenomenon may not be required for explaining the puzzling features of the universe which are giving rise to the inflation problem. I find it significant, however, that of all the major difficulties facing cosmology, the cosmological constant problem is the one for which inflation theory was never even considered to provide an appropriate answer, as this is definitely an issue that can only be addressed in the context of the generalized gravitational theory proposed in chapter 1. There may have been truth, then, to the long forgotten suggestion that the same insights that would turn out to be required in order to solve the cosmological constant problem may allow to do away with the other outstanding difficulties of cosmology which would otherwise need to be addressed by resorting to inflation theory.

Thus, while inflation may not be invalidated as a theory, it appears that

there are more natural solutions, based on more unavoidable theoretical constraints, not only to the inflation problem itself, but to certain other important issues as well. In fact, I'm now in position to provide satisfactory answers to practically all the remaining outstanding problems of theoretical cosmology including the problem of the origin of the arrow of time. But what should motivate one to recognize the necessity for an alternative approach to cosmology such as the one I have proposed in the preceding sections is the fact that even some of the originators of inflation theory have more recently expressed doubts concerning the usefulness of the theory for solving any of the problems to which it was originally believed to provide a satisfactory answer, because inflation itself requires very unlikely initial conditions to be initiated and to give rise to the desired outcome. Those criticisms, however, are usually overlooked because of what appears to be the overwhelming evidence in favor of inflation that is provided by the fact that the universe was confirmed to be flat and to be homogeneous above the scale of the horizon by observations of the cosmic microwave background.

Indeed, it is definitely the fact that a $\Omega = 1$ universe was always favored by inflation, even at a time when it appeared that lower values of Ω were favored by observations, that is responsible for having transformed inflation theory into the paradigm it is today, when it was later found that this parameter is in effect equal to unity and space actually is perfectly flat on the largest scale. But given that I have shown that in the presence of negative energy matter space must necessarily be flat based merely on the assumption that the universe must have zero energy, then it would appear that flatness is not valid as a confirmation of inflation, but is actually a generic property of any universe obeying the known principles of physics. In this context what was wrong was really the early expectation that by default (in the absence of inflation) space should be observed to be highly curved at the present time, given that perfect flatness appears to require very unlikely initial conditions. In fact, space must be perfectly flat at all times in a zero energy universe, but what I have tried to explain is that inflation has nothing to do with that and therefore flatness does not provide an unmistakable confirmation of the validity of inflation theory.

Of course if all I had done was to show that the flatness problem does not occur, even in the absence of inflation, when the universe is required to have a null energy, then it would not be possible to conclude that inflation is unnecessary, because there would still be a problem associated with the creation of matter and with the observed large scale homogeneity of the matter distri-

bution. But given that in the presence of negative energy matter and when one recognizes the necessity for all elementary particles in the universe to be causally related to one another, it actually becomes necessary for the initial matter distribution preceding the emission of the cosmic microwave background to be smooth enough that no macroscopic event horizon is allowed to be present, then it follows that the overall homogeneity of the temperature of CMB radiation is no longer a fact in need of some explanation.

The creation of all matter from nothing, on the other hand, no longer requires a hypothetical, post-inflation reheating process dependent on very specific conditions, as it actually is a basic requirement of a cosmological model involving negative energy matter that is naturally satisfied by the existence of pair creation processes involving opposite action particles. Indeed, such processes naturally allow matter to be created out of truly nothing, while the particles so produced can avoid annihilating back to nothing when expansion is sufficiently rapid, as it must have been in the first instants of the Big Bang in the context where the global expansion rate was then fixed to its critical value and the positive kinetic energy of expansion was required to compensate the enormous negative gravitational potential energy associated with the density of matter energy characteristic of quantum gravitational phenomena. When all the dust has settled, it appears that in fact not much evidence remains to possibly confirm that inflation really occurred. But again, that does not mean that none of the theoretical motivations behind inflation were justified, merely that inflation is not required to produce the apparently unlikely 'initial' conditions which were previously thought to be unexplainable outside the realm of this theory.

Concerning the flatness problem, however, it transpires that if the *specific* expansion rates of positive and negative energy matter were fixed to their critical value by inflation alone, while only the initial, average densities of positive and negative matter energy were required to be equal by the requirement of null energy (so that the gravitational potential energies and the kinetic energies of expansion were left unconstrained by the same condition), then it would be difficult to explain how the average, specific densities of the two opposite energy matter distributions could remain mostly the same following inflation, as required if the cosmological constant is to not be much larger than it currently is. There is no reason in effect to assume that the outcome of inflation would be exactly the same from both the viewpoint of positive energy observers and that of negative energy observers, while this is required if the specific densities are to be of similar magnitude following

inflation and therefore also at the present epoch. Thus, it would appear that the explanation of flatness provided in section 3.5 is actually an absolute requirement under such conditions and cannot merely be considered an alternative possibility.

But the most serious difficulty for an explanation of flatness through inflation is the fact that the theory relies on the hypothesis that a large non-vanishing value of vacuum energy density existed in the initial Big Bang state, which means that the process would actually have opposite effects of significant magnitudes on the expansion rates measured by observers with opposite energy signs. Thus, while the space experienced by a positive energy observer could be driven to inflate exponentially, the space experienced by a negative energy observer may actually be made to collapse back into a singularity, which means that the 'initial' densities of positive and negative energy matter measured by such an observer following inflation would remain maximum, while those measured by a positive energy observer would become minimum, which once again is not quite compatible with observations which indicate that the expansion rates and the spatial volumes experienced by opposite energy observers are still similar at the present epoch, due precisely to the very small value of the cosmological constant. It would therefore appear that additional fine tuning of a kind that hasn't even been considered yet would be required to make inflation theory viable.

What must be clear is that there is only one (positive or negative) value for the energy density of the vacuum at any given time and if the magnitude of this value is too large for too long a period of time initially, then there may be conflict with observations, even independently from whether matter is created initially (as I'm assuming) or after inflation (as must be assumed from a traditional viewpoint). In the context of the approach I favor this problem does not exit, because for a zero energy universe, when it is recognized that negative energy matter must be present in the initial state, there is a requirement for space to be perfectly flat from both the viewpoint of positive energy observers and that of negative energy observers and under such conditions, even if a non-zero average density of vacuum energy may be required to compensate any difference between the average density of positive matter energy and that of negative matter energy, the anthropic principle provides sufficiently strong a constraint to allow one to expect that the cosmological constant should nevertheless be as small as it is currently observed to be.

Concerning the solution potentially offered by inflation theory to the hori-

zon problem and more specifically to the unexplained uniformity of the initial matter distribution, it was already pointed out by Roger Penrose that the usual assumption to the effect that inflation would take the universe from a highly inhomogeneous state to a perfectly smooth one appears doubtful in the context where the initial state would in effect be characterized by a maximum gravitational entropy. But those remarks were made before we even had a theory of gravitation that allowed for the presence of negative energy matter in the initial state. From the viewpoint of the developments introduced in chapter 1 it would seem that it is definitely impossible to assume that a universe with an arbitrarily large gravitational entropy could be rendered homogeneous through accelerated expansion, as the potential for ever more dissociated positive and negative energy matter distributions is unlimited, just like the amount of matter itself. The opposite energy black holes that could be present in the initial state if it was not for the limitation exerted by the constraint of global entanglement on density fluctuations could be as massive as the radius of curvature of the universe is large and would concentrate all the matter in the universe in their gravitationally repelling singularities, which means that no amount of expansion could ever result in a homogeneous matter distribution. Thus, if negative energy matter does exist, it seems that inflation alone could not prevent the initial distribution of matter energy from being highly inhomogeneous and this provides additional motive to believe that the process is not necessary, even if it still cannot be ruled out that it might have occurred.

One particular aspect of the horizon problem which is currently believed to have been solved by inflation theory has to do with predicting the spectrum of density fluctuations in the initial positive energy matter distribution. Indeed, the fact that observations of the cosmic microwave background allow to deduce a scale-invariant spectrum of density fluctuations of the kind that is predicted by inflation theory is often considered to provide the strongest empirical evidence of the validity of this theory. But given that such a spectrum of density fluctuations would be typical of any theory according to which space itself does not have a characteristic scale, as is the case of any universe with a critical energy density and an infinitely large radius of curvature, then it seems that a well-behaved cosmological model that would describe a universe with null energy would also predict a scale-invariant spectrum of density fluctuations, given that it would necessarily predict that space is flat on a global scale and therefore, once again, the observations cannot be assumed to provide a definitive confirmation of inflation theory.

In any case I think that the fact that it was argued by certain authors that even if the gravitational wave signal attributable to inflation that we expect to eventually detect in the spectrum of CMB temperature fluctuations was not observed it would not necessarily mean that inflation theory is wrong, constitutes a perfect example of the unfalsifiable character which has come to define this theory and which makes it doubtful that it could ever be proven right. In the present case I would clearly be opposed to crossing that line and if the gravitational wave signal is not observed, then I believe that we should draw the necessary conclusion and reject the theory once and for all, because in such a case, even if we may still be unable to reject the possibility that an early phase of accelerated expansion occurred, this would no longer constitute the most viable solution to any particular aspect of the inflation problem.

Now, it has been hailed that the fact that certain inflation theories may allow the 'true' universe to be comprised of many regions like the known universe, separated by arbitrarily large portions of inflating space in which new 'universes' like our own are born all the time, could be a positive development given that it seems increasingly more likely that some properties of our universe are constrained by the anthropic principle. Indeed, one of the implications of the existence of such otherwise unexplainable properties is that it makes plausible the idea that there must be more than one possible instance of physical reality, so that the anthropically constrained universe we observe can exist as a mere possibility whose improbable nature need not be explained by appealing to divine intervention. Some of us, however, appear to favor, for some mysterious reason, that all of those realities instead of just existing on their own be somehow tied (however loosely) to the universe we do experience, as if this was a requirement of the multiverse concept. This conception of the multiverse has been appropriately renamed the 'megaverse' by Leonard Susskind and now enjoys respectable status as if it had been proved right by the 'successes' of inflation theory. But given that we may now have to reconsider the degree of inevitability of the phenomenon of inflation it would appear that all this extraneous amount of inflating vacuum may no longer be as appealing as it once was.

In any case, one thing should be clear and it is that eternal inflation is not necessary for making the multiverse concept a viable notion and in fact the emergence of a megaverse concept in inflation theory may actually constitute a problem for this approach to cosmology given that it may indefinitely postpone the moment at which the global entanglement of this whole enlarged

universe would occur in the past, while this must be considered a necessity, as I explained at length in the preceding section (if global entanglement is required only within the bubble universes then the megaverse itself could not be assumed to exist as an ensemble of causally related parts in the first place). This remark is all the more relevant in the context where, unlike the megaverse, the existence of an arrow of time (understood as being a consequence of global entanglement) is an observable fact with undeniably real consequences.

The most enduring problem facing inflation theory, however, remains the fact that it is still as difficult today as it was back when the model was introduced decades ago to identify what is the deep principle from which it would emerge as an unavoidable aspect of physical reality. If such a foundation cannot be developed we will perhaps eventually need to recognize that what was provided by inflation was a solution that was useful merely because of the absence of a better alternative. It is not appropriate in the context where a better explanation of facts is available to just keep adjusting the free parameters of a theory which is supposed to determine the very boundary conditions applying to the universe as a whole. At this point it is not just questionable whether inflation can actually solve any of the outstanding problems of Big Bang cosmology, it is even uncertain whether it is still possible to assume that the phenomenon occurred at all. Under such circumstances only our inherent resistance to paradigm change may prevent us from acknowledging the eventual failure of the theory. But if there is any reason to believe that inflation, in effect, did not occur, it would have to be the fact that it is not merely a single one of the difficulties originally assumed to be solved by the theory that can be explained away in the context of negative energy matter cosmology, but nearly all aspects of what was once the inflation problem.

3.11   Summary

To conclude this chapter I would like to provide a summary of the most significant results which were obtained concerning both the problem of dark energies and what I call the inflation problem in the context of the progress previously achieved in better understanding the concepts of negative energy matter, time reversal invariance, and black hole entropy. The following list offers an exhaustive account of those results. Once again the reader who may want to skip this section can do so without missing any essential developments

which were not already discussed in the present report.

1. While only the positive energy portion of a homogeneously distributed material substance with pressure opposite its energy sign like quintessence would influence the expansion rate measured by a positive energy observer, it is the observer independent sum of all positive and negative contributions to the density of vacuum energy which has an effect on the expansion rates experienced by positive and negative energy observers.

2. The vacuum fluctuation processes directly experienced by negative energy observers necessarily contribute positive and negative energies that are the exact opposite of those contributed by the vacuum fluctuation processes which are directly experienced by positive energy observers, because there must be a symmetry under exchange of positive and negative energy matter. But given that both categories of contributions exert a gravitational influence on matter of any energy sign then the natural value of vacuum energy density which we should expect to observe is zero, independently from how the symmetries of the chosen grand unified theory are broken in the current low energy state.

3. The maximum value of the contributions to vacuum energy density which can be directly measured by a negative energy observer is an invariant negative quantity which according to the requirement of symmetry under exchange of positive and negative energy states should be the exact opposite of the maximum value of the contributions to vacuum energy density which can be directly measured by a positive energy observer.

4. Any non-zero value of the cosmological constant measured by a positive energy observer would have to arise from a difference between the maximum positive and negative contributions to vacuum energy density provided by the natural vacuum-stress-energy tensors entering the generalized gravitational field equations associated with such an observer, in the context where the negative contribution must be submitted to the same kind of transformation as apply to the measures of negative energy matter density effected by a positive energy observer, which are not necessarily the same as the measures of negative energy matter density determined by a negative energy observer (even on a

global scale) in the context where observers of opposite energy signs do not share a unique perception of the metric properties of space.

5. Even when they are required to be very similar initially, the scale factors associated with the metric properties of space experienced by opposite energy observers can be made to differ significantly as a consequence of a difference in the expansion rates measured by opposite energy observers and given that the observed average value of vacuum energy density depends on such differences, then it follows that it should vary when the rates of expansion measured by opposite energy observers begin to differ.

6. When the positive cosmological constant is considered to be a manifestation of the non-vanishing measure of vacuum energy density, then the fact that the same positive vacuum energy modifies the rates of expansion experienced by positive and negative energy observers in opposite ways means that the near equality of those expansion rates which initially existed as a consequence of the constraints imposed by the weak anthropic principle and the condition of null energy constitute an unstable configuration which must give rise to a growing cosmological 'constant'.

7. If the cosmological constant and the density of vacuum energy are now positive despite the fact that the scale factor must have had nearly the same value from the viewpoints of opposite energy observers in the first instants of the Big Bang, then it means that based on the metric properties of space experienced by positive energy observers the universe must have expanded at a rate slightly higher than that which would be determined based on the metric properties of space associated with negative energy observers, which is appropriate given that the same vacuum energy exerts an opposite gravitational force on positive and negative energy matter.

8. A positive cosmological constant produces an acceleration of the rate of expansion of space measured by a positive energy observer and not merely an acceleration of the rate of expansion of positive energy matter, because the same metric transformation that is involved in determining the net value of vacuum energy density also affects the measures

of density of negative energy matter effected by a positive energy observer, which means that the initial ratio of positive to negative energy matter densities remains unchanged as space expands despite the fact that the *specific* density of negative energy matter (that which is measured by a negative energy observer) actually becomes larger than the *specific* density of positive energy matter (that which is measured by a positive energy observer) when the cosmological constant is positive.

9. The transformation of the metric properties of space that gives rise to a non-zero cosmological constant is made necessary merely as a consequence of the impossibility to directly compare the volume of a region of space experienced by negative energy observers with the volume of the same region experienced by positive energy observers, due to the fact that only the matter with a given sign of energy exerts an influence on the rate of expansion experienced by an observer with the same sign of energy.

10. If we were to assume that the sum of all contributions to vacuum energy which are directly experienced by positive energy observers add up to a maximum *negative* value, then a different form of generalized gravitational field equations would need to be considered such that from the viewpoint of a positive energy observer the metric conversion factor would apply to the positive portion of the maximum contribution to the density of vacuum energy, instead of applying to the negative portion of it. What justifies the validity of the original version of the equations is the fact that from the alternative viewpoint discussed here it would be difficult to reconcile the currently observed densities of positive energy matter and positive vacuum energy with the fact that the cosmological constant would rather tend to attenuate any divergence that would develop between the specific densities of positive and negative energy matter.

11. The smallness of the current value of the cosmological constant can only be explained by invoking the weak anthropic principle, as becomes possible in the context where we impose a requirement of null energy on the universe as a whole, because under such conditions the average density of vacuum energy is the only physical parameter that is allowed to vary despite the flatness of space and the null value of the energy contained in the gravitational field on the cosmological scale.

12. The presence in a uniform distribution of invisible negative energy matter of an underdensity attributable to the gravitational repulsion exerted on this matter by the presence of a visible positive energy matter overdensity would have the same effect on the surrounding positive energy matter as would the presence of an equivalent amount of gravitationally attractive dark matter.

13. In the context where there must be symmetry under exchange of positive and negative energy matter it is not possible to assume that there exists negative energy dark matter (dark from the viewpoint of both positive energy observers and ordinary negative energy observers) without assuming that there also exists positive energy dark matter with a similar but opposite average density and therefore it is not possible to attribute all the missing mass effects observed around positive energy structures to the presence of underdensities in the negative energy matter distribution.

14. The amplitude of missing mass effects attributable to the presence of underdensities in the negative energy matter distribution is limited by the finite value of the average density of negative energy matter and therefore the presence of negative energy matter underdensities can be expected to have accelerated the process of structure formation in the positive energy matter distribution only at the epoch in the remote past when the average density of matter was still relatively large and the matter was homogeneously distributed on the scale of the structures considered.

15. The additional gravitational attraction attributable to the presence of underdensities in the negative energy matter distribution must have accelerated the formation of the first galaxies in the positive energy matter distribution, which may allow dark matter to not be as susceptible to clumping as it would if it was composed of weakly interacting massive particles of the kind that is usually considered, which means that it cannot be excluded that the phenomenon behind dark matter does not really involve such particles.

16. Past the point at which the gravitational repulsion of a positive energy structure would be strong enough to give rise to a complete absence

of negative energy matter inside the structure, the missing mass effects attributable to a local reduction in the density of negative energy matter would reach a plateau and would no longer exert a significant influence on the gravitational dynamics of the visible positive energy matter.

17. The validity of the choice which is normally made to exclude the possibility that positive action electrons could exist which according to ordinary conventions would propagate a *positive* charge along with a negative energy backward in time can only be decided based on empirical evidence.

18. It is possible to assume that a certain portion of positive energy dark matter actually consists of the same particles as compose ordinary matter, but with opposite bidirectional charge signs (the signs of charge which are independent from the direction of propagation in time) when we acknowledge that those particles cannot directly interact with ordinary matter particles and anti-particles or transform into such particles on a continuous world-line, even though they do interact strongly among themselves, unlike more conventional dark matter particles.

19. The absence of interactions between ordinary particles and antiparticles and particles propagating reversed bidirectional charges forward or backward in time can be explained when we extend the requirement of continuity of the flow of time to interaction bosons by assuming that they propagate charges in both the forward and the backward directions of time all at once and then require that the bidirectional sign of charge remains normal (retains the sign of bidirectional charge which is normally attributed to known particles of the kind involved) along the direction of the flow of time associated with the world-lines of the interacting particles.

20. Given that the condition of continuity of the flow of time is a condition that applies merely to non-gravitational attributes, it cannot prevent a positive action particle with a given bidirectional charge sign to interact via the truly neutral gravitational interaction with another positive action particle carrying an opposite bidirectional charge, which means that such particles do exert attractive gravitational forces on one another and also repulsive gravitational forces on all negative action par-

ticles (because all negative action particles can be described as voids in the homogeneous distribution of positive vacuum energy and such voids always exert gravitational forces on positive action particles).

21. The magnitude of the average density of negative energy matter particles with reversed bidirectional charges is only required to be equal to the magnitude of the average density of positive energy matter particles with similarly reversed bidirectional charges for the same reason that justifies assuming that the magnitude of the average density of ordinary negative energy matter particles is equal to that of ordinary positive energy matter particles and therefore the actual density of positive energy matter particles with reversed bidirectional charges may differ from that of ordinary positive energy matter particles.

22. It is not possible to conclude that positive energy dark matter is composed entirely of baryonic particles with reversed bidirectional charges, because such particles would interact strongly among themselves and this would be incompatible with certain astronomical observations, given that it cannot be assumed that most of the matter has collapsed into massive compact astronomical objects at an early epoch.

23. The existence of local variations in the density of vacuum energy attributable to the differences that may emerge between the metric properties of space experienced by positive energy observers and those experienced by negative energy observers in the presence of local gravitational fields produced by inhomogeneities in the matter distribution, would contribute to increase the mass of large astronomical objects in a way that would allow to reproduce most of the observed missing mass effect around visible large scale structures, given that such concentrations of vacuum energy would themselves generate local gravitational fields that would give rise to further local variations of vacuum energy density.

24. Despite the fact that dark matter, for the most part, actually constitutes a form of vacuum energy, it nevertheless contributes like ordinary positive or negative energy matter to the average density of energy on the cosmological scale, because its presence is attributable to *local* variations of vacuum energy density, which means that it must be

assimilated with the presence of voids in the otherwise uniform distribution of vacuum energy whenever its energy is opposite the energy of the observer which is experiencing its gravitational field, while a uniform distribution of such underdensities exerts no gravitational force on matter with an opposite energy sign. As a result, as long as it is homogeneously distributed on a global scale, negative energy dark matter does not contribute to decelerate the expansion rate determined by positive energy observers, which means that the average densities of positive and negative matter and vacuum energy are allowed to reach their theoretically and observationally required critical values (because there is no compensation of positive energy contributions by negative energy contributions on the cosmological scale, unlike is the case with ordinary positive and negative contributions to the density of vacuum energy).

25. In order to avoid contradictions it is necessary to assume that despite the fact that the portion of missing mass effects attributable to local variations in the density of vacuum energy can be expected to grow along with the inhomogeneity of the large scale matter distribution, the average densities of positive and negative energy dark matter do not change with time, which means that the amount of energy that cannot be accounted for by the ordinary matter present in a large astronomical structure already existed in diffuse form before the formation of that structure, despite the fact that it was not exerting a detectable gravitational force locally.

26. The observed strong correlation between the gravitational acceleration attributable to the total amount of matter inside an orbit and the gravitational acceleration attributable to the normal matter confirms that the missing mass effect is attributable to local variations in vacuum energy density, because in such a context the presence of dark matter must be considered to be an effect of the curvature of space attributable to the matter that is present in a region of space on the local measures of vacuum energy density and the more gravitational acceleration there is as a consequence of the presence of normal matter, the more distinct the metric properties of space experienced by opposite energy observers must be and therefore the more dark matter there must be.

27. An explanation of the missing mass effect as being a consequence of

local variations in the density of vacuum energy is preferable to an explanation based on a modification of the laws that govern the gravitational dynamics of astronomical objects, because it allows the portion of dark matter that arises from the presence of vacuum energy inhomogeneities to clump like ordinary dark matter, which means that once created a dark matter inhomogeneity can persist all by itself for some time, sustained by its own gravitational field, even when it becomes separated from the ordinary matter that gave rise to it, as is observed to happen in the course of certain collisions involving large astronomical objects.

28. The information concerning the signs of bidirectional charge is not lost when matter is captured by the gravitational field of a black hole, because the information about the direction of propagation in time of matter particles with reversed bidirectional charges (from which depends the signs of their charges determined from the unidirectional time viewpoint) is contained in the microscopic state of a component of the electromagnetic field that is distinct from that with which ordinary matter particles (with normal bidirectional charge signs) interact.

29. In the context of the proposed interpretation of reversed bidirectional charge particles, four parameters are allowed to vary for a particle under the influence of an elementary black hole with a given energy sign, which are the direction of its momentum (which varies as a function of the sign of energy of the particle itself), its handedness, its direction of propagation in time, and the sign of its unified bidirectional charge. Four truly elementary units of area (each equal to a Planck surface) are therefore required to encode the information about the exact microscopic state of the particle on the surface of the black hole. But given that only the information concerning the handedness of the particles contained in a macroscopic stable state black hole needs to be encoded in the microscopic state of the gravitational field on the surface of the object it follows that when we consider only the thermodynamic properties associated with the surface gravitational field of such a black hole we obtain a value for its entropy that is equal to only one fourth the area of its event horizon measured in such truly elementary units of surface.

30. As a consequence of the presence of negative energy matter in our

universe there exists an additional source of gravitational instability which arises not from stronger gravitational attraction, but from the gravitational repulsion exerted on visible matter by negative energy matter galaxies and clusters.

31. Given that in the presence of negative energy matter the inhomogeneities present in the positive and negative energy matter distributions contribute to reinforce one another, then it can be expected that under such conditions the rate of development of large scale structures is accelerated, which allows to more easily reconcile the high level of development of present day inhomogeneities with the near perfect uniformity of the temperature of cosmic microwave background radiation.

32. The additional contribution to the gravitational field of a void in the positive energy matter distribution which arises from the presence of an invisible negative energy matter overdensity inside this void would allow smaller voids to exert an unexpectedly large gravitational repulsion and in such a way allow to explain the observations which suggest that certain voids in the matter distribution do exert larger than expected gravitational repulsion on galaxies located in their periphery.

33. When an overdensity of negative energy matter is present inside the voids in the visible matter distribution it is possible to explain observations which show that there is a much smaller number of galaxies in the Local Void than is predicted by computer simulations, because any galaxy that would form in the void would rapidly be expelled to the periphery by larger than expected repulsive gravitational forces.

34. Given that the density of negative energy matter in the Local Void can be expected to be higher than it is in our galactic neighborhood, it follows that the missing mass effects attributable to negative energy matter underdensities would be more localized around those galaxies located nearer the void and this would have accelerated the formation of positive energy galaxies in this area, which allows to explain why a larger than expected number of very large galaxies in the Local Sheet are located on the periphery of the Local Void.

35. The existence of certain bulk flows could be explained as being the consequence of the presence of a very large scale polarization of the

primordial distribution of positive and negative matter energy, which is not forbidden by the constraint responsible for the overall homogeneity of the primordial matter distribution and which would give rise to an alignment in the fluctuations of CMB temperature as a result of the fact that such an inhomogeneity would produce a very large scale variation in the gravitational potential which would contribute to further enhance the polarization of the primordial matter distribution by creating a force field that would accelerate positive and negative energy galaxies in opposite directions.

36. In the context where the presence of negative energy matter is equivalent to missing positive vacuum energy it follows that the magnitude of the fine-structure constant α could vary in space or in time as a consequence of the fact that this coupling constant is affected by the virtual processes taking place in the vacuum, which means that if energy is missing from the vacuum that would normally be carried by the virtual particles which directly interact with positive energy matter, then the renormalized value of the fine-structure constant could be reduced or increased in proportion to the amount of energy that is missing, which is proportional to the amount of negative energy matter that is present.

37. The phenomenon of repulsive gravitational lensing which should occur when the visible light from a distant source is gravitationally repelled while it travels through a negative energy matter overdensity or a positive energy matter underdensity before reaching our telescopes would distort the image of background structures in such a way that the objects observed would appear to be more densely packed in space, possibly producing blobs of light and this could explain some observations which appear to show the presence of unexpectedly large quasar groups in the very distant past.

38. In a quantum gravitational context a limit exists to the magnitude of the positive and negative contributions to the density of matter and vacuum energy which is determined by the natural vacuum-stress-energy tensors associated with the upper limits of the positive and negative contributions to vacuum energy density and therefore space cannot continue to contract in the past beyond the point at which this limit is reached.

39. The total energy of the universe as a whole is a physical property which must necessarily be defined in a purely relational way and therefore it must have a null value, because if that was not the case then this value would have to be either positive or negative and this would allow the particular direction of time relative to which this positive or negative energy would propagate to be singled out as an absolutely defined direction.

40. When the average, specific density of negative energy matter is growing relative to that of positive energy matter as a consequence of the emergence of a difference between their specific rates of expansion, the ratio of the average densities of positive and negative energy matter determined by a positive energy observer remains invariant, because the density of negative energy matter measured by such an observer is modified by the same metric conversion factor which fixes the density of vacuum energy, while the density of vacuum energy grows in proportion to the magnitude of the divergence between the scale factors experienced by opposite energy observers. As a result, if the total density of matter energy had been null in the initial Big Bang state then it would remain so as expansion takes place, in accordance with the requirement of relational definition of physical attributes.

41. If it was not for the constraint imposed by the condition of zero energy there would be no a priori motive to assume that the gravitational potential energy associated with the curvature of space on the cosmological scale should be the same for positive and negative energy observers at the same epoch, because the kinetic energy of expansion determined by an observer with a given energy sign would not be required to compensate the difference between the positive and the negative energy of matter, while the expansion rate that determines the kinetic energy of expansion is an observer dependent quantity that need not necessarily be the same for observers with opposite energy signs.

42. The initial value equation derived from the gravitational field equations merely expresses the requirement of gravitational energy conservation for the universe as a whole, but it does not, all by itself, require that the universe comes into existence with zero gravitational energy in the context where it is recognized that the real measure of gravitational energy for the universe is that which is associated with the curvature

of space on the cosmological scale. Under such conditions what allows the rate of expansion to have a rather large and critical value and the gravitational energy associated with the curvature of space to have a null value for an expanding zero energy universe is the fact that the gravitational potential energy of the matter that must balance the kinetic energy of expansion experienced by a positive energy observer can be arbitrarily large even when negative energy matter is present and the total energy of matter itself is null.

43. Even if the magnitude of positive matter energy may be equal to that of its associated negative gravitational potential energy in the initial Big Bang singularity it is not this gravitational potential energy that *must* balance the positive energy of matter itself, as earlier creation out of nothing proposals required assuming.

44. What is required by the condition of zero energy for the universe as a whole is merely that the measure of gravitational field energy associated with the initial value of the (redefined) spatial curvature parameter $-k/a^2$ determined using the metric properties of space experienced by positive energy observers necessarily be such that it balances any residual energy of matter obtained by adding the opposite contributions of positive and negative energy matter.

45. It can be expected that on a very short time scale pairs of opposite action particles are continuously being created out of nothing without violating the constraint of energy conservation and when it is required that the kinetic energy of expansion measured by a positive energy observer during the Big Bang balances the large negative gravitational potential energy of the positive energy matter present under such conditions, then the expansion rate can be sufficiently large over a sufficiently long period of time to allow those particles to avoid annihilating back to the vacuum as they normally would.

46. The creation of all matter out of nothing through opposite action pair creation processes is a necessary but not sufficient condition for obtaining an initial state where the total energy of matter would be exactly null on a global scale, because in principle, even for a zero energy universe, the energy of the positive action particles can still be larger or smaller than the energy of the negative action particles on the average,

as long as the differences between the positive and the negative energies of matter are compensated from the viewpoint of a given observer by the energy of the gravitational field associated with the curvature of space, which depends on the rate of expansion measured by that observer.

47. The nullity of the energy of matter cannot be fixed as an independent condition, because that would require assuming that no local fluctuations above or below the zero of energy of matter can be present in the initial Big Bang state (despite the fact that such variations in matter energy could be compensated even in a maximum density state by local variations in the kinetic energy of expansion above or below the value corresponding to a critical expansion rate), while the presence of such fluctuations is required to explain the observed inhomogeneities present in the initial distribution of matter energy on a scale larger than the cosmic horizon.

48. Even though it is necessary to assume that both the positive and the negative energy of matter contribute to the total measure of energy effected by a positive energy observer in the context where all matter is created out of nothing as opposite action pairs, it is nevertheless the case that only the energy of the gravitational field perceived by a positive energy observer can contribute to the energy budget that must add up to zero on a global scale from the viewpoint of such an observer, because the rate of expansion and the gravitational potential energy of matter which allow to determine this energy are observer dependent properties which are only significant from the viewpoint of the metric properties of space experienced by an observer with a specific sign of energy.

49. If the positive energy of matter was larger than its negative energy for a zero energy universe and vacuum energy provided a negligible contribution to the energy budget, then the gravitational energy associated with the curvature of space would need to be negative, but while this would mean that space is positively curved and closed from the viewpoint of a positive energy observer, it would also mean that space is negatively curved and open from the viewpoint of a negative energy observer and therefore it would appear that only a perfectly flat universe

with zero gravitational energy could be open from both the viewpoint of a positive energy observer and that of a negative energy observer.

50. Given that the difference between the volume of a closed universe and that of an open universe would in principle be infinite regardless of the exact curvature of their spaces, it may seem that if the universe was not perfectly flat the magnitude of the density of vacuum energy would already be maximum initially, but this just cannot be the case because vacuum energy also contributes to the positive or negative density of energy which must be canceled out by the energy of the gravitational field, so that when the average density of positive matter energy differs from that of negative matter energy, the non-zero density of vacuum energy simply compensates the non-zero energy of matter, thereby making space flat from the viewpoint of all observers.

51. If vacuum energy density was too large in the first instants of the Big Bang, then it would prevent the emergence of an observer and therefore the weak anthropic principle would not allow a configuration where a large disparity exists between the magnitude of the average initial density of positive matter energy and that of negative matter energy.

52. In a zero energy universe, regardless of the actual value of the cosmological constant, the kinetic energy of expansion determined by a positive energy observer must always precisely compensate the gravitational potential energy attributable to positive matter energy and vacuum energy, while the kinetic energy of expansion determined by a negative energy observer must always compensate the gravitational potential energy attributable to negative matter energy and the same vacuum energy (whatever its energy sign) and therefore the fundamental principle that allows to determine which solution of the gravitational field equations is the appropriate one for a description of the expansion of space on the cosmological scale is the requirement of relational definition of physical attributes applied to the energy of the universe as a whole.

53. Observations appear to confirm the validity of the conclusion that the magnitudes of the average densities of positive and negative matter energy must be very similar in the initial Big Bang state, because the relatively low amplitude of fluctuations in the temperature of the cosmic

microwave background implies that the magnitude of initial inhomogeneities in the distribution of negative matter energy is comparable to that of the inhomogeneities present at the same epoch in the distribution of positive matter energy, while the magnitude of the inhomogeneities present in any distribution of matter energy at the epoch of recombination depends on the average matter density.

54. If the processes of matter creation out of nothing that took place initially are no longer occurring long after the Big Bang, despite the fact that the densities of positive and negative energy present in the vacuum remain as large as they initially were, it is because once real matter is created as a consequence of the rapid initial rate of expansion, only positive energy matter influences the rate of expansion measured by a positive energy observer and this means that this expansion rate rapidly decelerates to the point where matter can no longer be permanently created.

55. The initial push of inflation is not necessary to explain the fact that the universe is actually expanding, because in the context where a maximum density of opposite action particle pairs must be assumed to be fluctuating out of nothing at every moment in the vacuum, if the gravitational potential energy of positive matter and vacuum energy must be entirely compensated by the kinetic energy of expansion that is measured by a positive energy observer initially, as allowed when the total energy of matter and vacuum is null, then expansion must take place at a rate that is sufficiently large for expansion to persist over an arbitrarily long period of time.

56. Even if time was assumed to be continued past the Big Bang singularity following a hypothetical Big Bounce, processes of opposite action pair creation out of nothing would still need to be responsible for the existence of all the matter in the universe, because the 'final' state which would be reached while space collapses in the future direction of time prior to the initial singularity would be identical from a macroscopic perspective to the state that provides our initial boundary conditions and under such conditions most of the matter that is already present would annihilate back to nothing.

57. When inflation is not required to occur, matter does not have be created

by a delayed process of reheating in order that the Big Bang in effect be hot.

58. When the average, specific density of positive matter and vacuum energy and the average, specific density of negative energy matter are both fixed very precisely to their critical value initially, the current positive value of the cosmological constant implies that the magnitude of the average, specific density of negative energy matter is larger than the magnitude of the average, specific density of positive energy matter by an amount equal to the current value of vacuum energy density, because both the specific density of negative energy matter plus vacuum and the specific density of positive energy matter plus vacuum must have remained critical if they originally were.

59. Despite the fact that entropy is a measure of missing information, an objective characterization of temporal irreversibility does not require assuming that the information associated with the structures present on a microscopic scale is actually vanishing from reality when entropy is rising, because ignorance merely grows as a consequence of the fact that the macroscopic parameters we use to describe the state of a system are leaving aside an increasingly larger portion of the information that would be required to accurately describe its exact microscopic state.

60. In the context where the paths followed by the elementary particles forming a macroscopic system are fundamentally unpredictable it cannot be assumed that the entropy decreasing evolution which is continuously taking place in the past direction of time is the consequence of a precise adjustment of the present conditions that would predispose those systems to evolve in such a way when the state of motion of all particles is reversed.

61. The presence of event horizons provides us with a unique set of macroscopic physical parameters which allow a natural definition of coarse-graining and therefore an objective measure of entropy growth which cannot be attributed merely to the choice of macroscopic parameters (it is not the outcome of any *arbitrary* definition regarding what parameters should characterize the macroscopic state of the system), because when black holes are involved no other macroscopic parameter is available to define an alternative measure of entropy.

62. The growing amount of missing information that would be required to completely specify the state of all the microscopic, binary degrees of freedom on the surface of a black hole of increasing mass reflects the existence of a growing amount of structure in the microscopic state of the gravitational field which constitutes an objective property of the event horizon of such an object and which means that no information is lost when a black hole absorbs low entropy matter.

63. Once we recognize that an objective increase in the amount of missing information required to completely describe the detailed microscopic state of the gravitational field is taking place locally when stronger gravitational fields develop as a result of the formation of a matter overdensity, then it becomes necessary to admit that new degrees of freedom, which characterize the exact microscopic state of the gravitational field, are being created when the entropy of a black hole is growing, because the amount of missing information which would be required to specify the exact microscopic state of all the matter particles which were captured by the gravitational field of a black hole is not large enough to account for its entropy growth.

64. Despite the fact that locally the amount of missing information is growing faster than would appear to be allowed as a mere consequence of growing ignorance concerning the microscopic state of matter when gravitational fields gain in strength as a result of gravitational collapse, the total amount of information required to describe the exact microscopic state of our universe does not really change and the requirement of conservation of information is not violated, because an opposite variation of the same kind occurs when local gravitational fields grow stronger as a result of the formation of an underdensity of similar magnitude in the large scale matter distribution which allows a compensation between those two opposite variations to occur that leaves information invariant.

65. It is due to the fact that it is not necessary to specify the value of the microscopic gravitational degrees of freedom which are absent as a result of the reduced amount of gravitational interaction that is attributable to the presence of a local underdensity in the positive energy matter distribution that the microscopic state of the gravitational field can be specified using a smaller amount of information under such conditions.

66. Even if the amount of information necessary to describe the exact microscopic state of the gravitational field is diminishing locally when an underdensity is forming in the matter distribution, the total measure of gravitational entropy is still growing in the universe, because the amount of information necessary to describe the exact state of the gravitational field attributable to the matter overdensity that must form in the surrounding space as a result of the formation of this underdensity is growing and is now accounted for as missing information.

67. Even though some local gravitational fields experienced as repulsive by positive energy matter must be assumed to grow stronger merely as a result of the reduced level of gravitational interaction that is consequent to the formation of a void in the matter distribution, the changes involved are still thermodynamically favored, because they are always accompanied by the formation of matter overdensities which produce stronger attractive gravitational fields with which is associated a larger gravitational entropy.

68. Despite the fact that from an external viewpoint the gravitational field produced by the presence of an overdensity of *negative* energy matter appears to be equivalent to the gravitational field attributable to an underdensity in the *positive* energy matter distribution, a clear distinction exists between those two configurations with regards to thermodynamic properties, because whereas a negative energy black hole has negative temperature and radiates negative energy particles, a void of corresponding amplitude in the positive energy matter distribution would rather absorb positive energy particles as a consequence of having the same negative temperature.

69. Despite the similarity of the gravitational fields attributable to voids in a matter distribution and those attributable to overdensities of opposite energy sign, there exists a fundamental difference between those two categories of objects which arises from the fact that negative energy objects do not consist of voids in a positive energy matter distribution, but are rather equivalent to voids in the positive energy portion of the vacuum.

70. What explains that the formation of a void in a uniform positive energy matter distribution would give rise to a diminution in the amount of in-

formation required to describe the microscopic state of the gravitational field, while the formation of a void of similar magnitude in the positive energy portion of the vacuum would produce a positive change in the measure of missing information concerning the gravitational field, is the fact that in the absence of local variations of vacuum energy density associated with the presence of dark matter, the distribution of vacuum energy is really uniform on all scales, while the 'homogeneous' distribution of matter in which a void may be produced is not really uniform on a smaller scale, which means that a certain amount of structure is contained in its gravitational field that is lost when matter is removed, while removing energy from the vacuum rather produces additional structure in the gravitational field, given that it is equivalent to locally increasing the density of matter with an opposite energy sign.

71. Even the gravitational field attributable to a smooth matter distribution would contribute a certain measure of information which would be reduced as a result of the expansion of space and which would thereby compensate the growth in the amount of information that takes place as a result of the production of additional elementary units of space on the quantum gravitational scale which also arises as a result of the expansion of space and which would otherwise violate the rule of conservation of information.

72. As a result of the distinction that exists between the variation of gravitational field information arising from a local variation in the density of matter and any possible variation of gravitational field information which would be produced by a similar change in the average density of vacuum energy it follows that a variation of the average value of the cosmological constant would not contribute to alter the total amount of missing information contained in the microscopic state of the gravitational field, despite the fact that, like a variation of the average density of matter energy, a variation of the average density of vacuum energy would provide a variable contribution to the gravitational field that determines the rate of expansion. The fact that this is required for information to be conserved on a global scale constitutes a confirmation of the validity of the hypothesis that there actually exists an amount of structure in the gravitational field associated with a homogeneous mat-

ter distribution that is absent from a uniform distribution of vacuum energy.

73. The fact that the density of matter was much larger in the past does not make the initial smoothness of the matter distribution more unexpected, because when negative energy matter is present in the universe there may exist inhomogeneities on the largest scale even in a low density state, as a result of the fact that negative energy matter can be concentrated in regions of space distinct from those occupied by positive energy matter, which means that even if the initial state was assumed to be a low density state the most likely configuration for the initial distribution of matter energy might still be one of higher inhomogeneity and arbitrarily strong local gravitational fields with which would be associated a maximum measure of gravitational entropy.

74. The fact that the universe must not be collapsing locally if an observer is to be present to witness an absence of inhomogeneities does not provide strong enough a constraint to explain that the initial distribution of matter energy was as smooth as it is observed to be, even if the presence of event horizons would indeed prevent space from expanding locally, because matter energy could be much more inhomogeneously distributed than it currently is and expansion would still be allowed to proceed unaffected in most locations.

75. The homogeneity of the initial distribution of matter energy is not merely apparent in the low magnitude of local variations in the energy of elementary particles (which can be compensated by local variations in gravitational energy), but would also be apparent in the near absence of large scale disparities in the distribution of positive and negative energy matter particles.

76. Even in the presence of negative energy matter it is still appropriate to consider that there arises a state in the past which from a classical viewpoint would be characterized as consisting in a spacetime singularity, because if the distribution of negative energy matter is sufficiently homogeneous on the largest scale initially, it would exert no influence at all on the rate of expansion of positive energy matter, which means that it would not diminish the strength of the gravitational field attributable to the presence of this matter and therefore it could not

prevent the formation of the trapped surface which according to classical theorems would give rise to a past singularity, even if one of the axioms of the theorems is that matter must have positive energy.

77. Even if there exist solutions of the gravitational field equations that would appear to describe processes which would be the time-reverses of a black hole gravitational collapse, those hypothetical white holes are never observed, because they would require a decrease of gravitational entropy in the future, which is unlikely in the absence of a specific constraint, or equivalently a continuous increase of gravitational entropy in the past direction of time, which is forbidden in our universe.

78. When both positive and negative energy matter are present in the universe, the initial state which would be characterized by the highest gravitational entropy would be one for which the distributions of positive and negative energy matter would be completely dissociated in such a way that all the matter would be contained in opposite energy black holes with arbitrarily large masses whose magnitude would be limited solely by the amount of matter in the universe.

79. Inflation cannot explain the observed time asymmetry that characterizes cosmic evolution, because there is no reason to expect that a contracting universe would evolve toward a more homogeneous configuration during the epoch that would precede a hypothetical phase of exponentially accelerating contraction which would take the universe back to a more likely state of maximum inhomogeneity, unless the state immediately preceding the exponentially accelerated contraction into a final singularity was required to be as smooth as the state which was produced in the past following ordinary inflation, which would amount to require without motive that causality operates backward in time from the final singularity, while classical (unidirectional) causality, or the rule that past events have an influence on future events and not the opposite is simply a manifestation of the irreversibility of time and if this property must be assumed to characterize the evolution that takes place in one or another direction of time then it cannot be used to explain the time asymmetry itself.

80. The homogeneity of the 'initial' distribution of matter energy cannot be explained as being merely a consequence of the fact that the presence

of very large density fluctuations in the initial state of a creation out of nothing event would not allow matter to be produced out of nothing by processes of opposite action pair creation, because in the absence of an independent constraint concerning the homogeneity of the distribution of matter energy that emerged from the past singularity following a quantum bounce it is simply not possible to assume that there is a necessity for all matter to be created out of nothing during the Big Bang.

81. The notion that the size of the cosmic horizon increases with time is dependent on the implicit assumption that the classical principle of causality is valid and that effects propagate in the future direction of time from past causes which originate in the initial singularity and therefore the very validity of this horizon concept is dependent on the existence of a constraint regarding the homogeneity of the initial distribution of matter energy from which time asymmetry arises, which means that it is really a solution to the horizon problem that must be based on an explanation of the origin of time asymmetry and not the opposite.

82. When the size of the cosmic horizon reaches the limit in the past at which the non-locality that is intrinsic to quantum phenomena becomes prevalent, the limited velocity of causal signals no longer forbids the existence of causal relationships between the particles that could be present within the volume of the horizon and in fact it is not even possible to assume the existence of particles smaller than this natural scale of quantum gravitational phenomena, so that there would be no sense in imposing limitations on signal propagation below that scale.

83. If all the elementary particles originally present in the universe at the Big Bang and from which evolved the current matter distribution were allowed to be in contact with at least one other particle of any energy sign when the size of the cosmic horizon reached the quantum gravitational limit in the past, then no particle would exist in the universe that would be causally unrelated to the other particles, even as they become separated by large distances, which is sufficient a condition to allow the universe to form a globally consistent whole.

84. If only positive energy matter was present in our universe it seems

that the global entanglement constraint which imposes contact between neighboring elementary particles at the Planck time could be satisfied in the initial state without gravitational entropy being minimal (and macroscopic black holes being absent), but given that gravitational repulsion, unlike gravitational attraction, may forbid local contacts between opposite energy particles by giving rise to insurmountable potential barriers for particles located within opposite energy black hole singularities, it follows that when negative energy matter is present in the universe event horizons can be expected to be absent initially on all but the smallest scale.

85. The fact that an ensemble of systems started in the same macroscopic state evolve to occupy all available microscopic states in the future, while a similar ensemble started in the same macroscopic state usually evolve only to past states characterized by a lower gravitational entropy is a consequence of the necessity that all the elementary particles present in the maximum density state at the Big Bang be allowed to come into contact with their neighbors of any energy sign in order that there exist causal relationships between all independently evolving components of the universe and by its very nature this condition allows to explain the fact that it is only the gravitational component of entropy which was not already maximum at the Big Bang.

86. The condition of global entanglement only imposes an absence of macroscopic event horizons while the event horizons associated with the presence of elementary (quantum gravitational) black holes need not be absent, because such an object merely constitutes the surface of the one and only elementary particle whose motion it constrains, which means that particles which are under the influence of the gravitational fields of elementary black holes in the state of maximum matter density are still allowed to come into contact with one another regardless of their energy signs.

87. Gravitational entropy must be decreasing continuously in the past direction of time regardless of whether space is expanding or contracting, as long as we are indeed approaching the instant at which is formed the unique singularity on which the condition of global entanglement is to be imposed, because the condition that applies to the initial singularity is precisely one of minimum gravitational entropy from which must

necessarily emerge a phenomenon of classical (unidirectional) causality that operates toward the future from that particular instant of time at which the cosmic horizon begins to spread, thereby allowing to impose on the future that it be such as to give rise to this highly constrained initial state.

88. The constraint of global entanglement enables time asymmetry to be derived from fully time-symmetric physical laws, because it merely allows to determine the boundary conditions applying on the macroscopic state of the gravitational field at the Big Bang.

89. All physical systems, regardless of how isolated they may have become and independently from how carefully they are prepared, must evolve with continuously decreasing gravitational entropy in the same direction of time toward the initial singularity, because all systems in the universe are submitted to the same unavoidable constraint applying to this unique state of maximum matter density as a consequence of the requirement that they actually be part of the same universe and of no other.

90. The fact that the parallelism of the asymmetry of thermodynamic evolution is observed to apply under all circumstances shows the validity of the arguments that allowed me to determine the nature of the constraint that imposes such an asymmetric evolution even on isolated branch systems.

91. A state of maximum matter density must necessarily occur at one time or another for the global entanglement of all elementary particles to be satisfied and given that such a state would not likely be characterized by an absence of macroscopic event horizons unless it constitutes the mandatory unique event at which global entanglement is enforced on the universe, then one must conclude that our Big Bang really is this unique event.

92. It is the fact that the condition of global entanglement would only be required to apply once, even if the universe was to return to a state of maximum matter density at some point in the future, that explains that the evolution that takes place from the moment at which this condition is enforced is not symmetric in time.

93. If time extends past the initial singularity following a quantum bounce, space will be expanding and the density of matter will be decreasing immediately after the bounce, while the inhomogeneity of the matter distribution will still be minimum, which means that the thermodynamic arrow of time will initially have the same direction as the cosmological arrow of time associated with expansion and will actually be opposite that we observe on our side in time of the initial singularity, which would make the universe completely symmetric under exchange of past and future.

94. In the context where gravitational entropy is growing along with the degree of dissociation of the positive and negative energy matter distributions, if there is an infinite amount of matter in the universe, then gravitational entropy may be allowed to rise indefinitely.

95. When the initial distribution of matter energy is uniform to a very high degree (as required for global entanglement to take place) and the local rates of expansion of positive and negative energy matter only vary as a function of the difference between the local amplitudes of their energy density (as required for energy to be null), then the rate of expansion of positive energy matter itself does not vary much from place to place and the matter distribution remains homogeneous and isotropic.

96. As a result of the uniformity of the initial expansion rate that follows from the constraint of global entanglement and the requirement of null energy for the universe as a whole, the cosmic microwave background can be expected to be homogeneous even on a scale larger than the size of the cosmic horizon at the epoch of recombination, because the absence of macroscopic event horizons is required on all scales, and this also means that no independent assumption is required to confirm the relevance of the cosmological principle for a description of the early universe on the largest scale.

97. The constraint of global entanglement does not impose a perfect homogeneity on the initial distribution of matter energy, but merely imposes an upper bound on the amplitude of fluctuations in the density of matter, while the principle of conservation of energy and the requirement of relational definition of physical attributes only require the universe to be flat and the energy of matter and vacuum to be null on the scale

of the universe as a whole and does not forbid the density of negative matter to differ from that of positive energy matter locally and therefore small inhomogeneities in the distributions of positive and negative matter energy should exist on all scales in the initial state that emerges from the Big Bang singularity which means that no independent solution is required for the smoothness problem.

98. The cause of the existence of correlations between the density fluctuations present in the initial distribution of matter energy on a scale larger than the size of the cosmic horizon is the global entanglement constraint itself, which also requires a certain local smoothness in the inhomogeneities which are allowed to be present initially, thereby giving rise to the presence of structures above the horizon size, which need not have been produced by the propagation of causal influences.

99. The relatively low abundance of topological defects may simply be a consequence of the fact that the presence of such high energy objects in the initial Big Bang state is incompatible with the requirement of smoothness of the primordial distribution of matter energy that is imposed by the constraint of global entanglement, even if merely in the sense that the amplitude of initial fluctuations in the energy density of matter may be too small to allow the production of highly dense topological defects at later times.

100. Of all the measurements concerning the spectrum of temperature fluctuations in the cosmic microwave background only those that concern a determination of the angular size of fluctuations (from which are derived the average density of positive matter and vacuum energy) would remain mostly unaffected by the presence of negative energy matter, because there are no influences arising from the presence of a homogeneous distribution of negative energy matter on the trajectories of positive energy photons from which is determined the geometry of the universe.

101. We will not observe the gravitational wave signal which according to traditional models should show up in the polarization of CMB radiation as a consequence of the stretching of primordial density fluctuations by inflation, because even if such a process took place, gravitational waves would not be produced abundantly given that the primordial

matter distribution must have already been highly homogeneous before inflation occurred.

102. Given that in the presence of negative energy matter space must necessarily be flat if the universe is to have null energy, then it would appear that the property of flatness is not valid as a confirmation of inflation, but is actually a generic characteristic of any universe obeying the known principles of physics.

103. Given that in the presence of negative energy matter and when one recognizes the necessity for all elementary particles in the universe to be causally related to one another, it becomes necessary for the initial matter distribution to be uniform enough that no macroscopic event horizon is allowed to be present, then it follows that the near homogeneity of the temperature of CMB radiation does not constitute a proof that inflation occurred.

104. The creation of all matter from nothing no longer requires that some post-inflation reheating occurs, as it can be naturally satisfied by the existence of pair creation processes involving opposite action particles in the context where the initial rate of expansion is required by the condition of zero energy to be large enough over a long enough period of time that it allows the particles so produced to avoid annihilating back to nothing.

105. If the specific expansion rates of positive and negative energy matter were fixed to their critical value by inflation alone, while only the initial, average densities of positive and negative matter energy were required to be equal by the requirement of null energy (so that the gravitational potential energies and the kinetic energies of expansion were left unconstrained by the same condition), then it would be difficult to explain how the average, specific densities of the two opposite energy matter distributions could remain mostly the same following inflation, as required if the current magnitude of the cosmological constant is to not be much larger than allowed by observational constraints.

106. In the presence of negative energy matter it becomes impossible for a universe with an arbitrarily large gravitational entropy to be rendered homogeneous through inflation, as there is no limit to the degree of

dissociation of the positive and negative energy matter distributions and if it was not for the limitation exerted by the constraint of global entanglement on density fluctuations, opposite energy black holes could be present in the initial state that would be as massive as the radius of curvature of the universe is large.

107. The megaverse of eternal inflation is not necessary for making the multiverse concept a viable notion.

Conclusion

Main results

I have come a long way since first asking what would happen to a negative mass object dropped in the gravitational field of the Earth. Yet I was able to confirm that my early intuition was right and that consistency dictates that the negative mass would need to 'fall' upward despite the fact that this goes against current expectations. This is a conclusion for which I have provided ample justification and even if that was all I had been able to establish I would already be very satisfied with the outcome of my undertaking. But several other developments were introduced in this report which are all related to the issue of time directionality as a concept independent from the thermodynamic arrow of time. In fact, the hypothesis of the existence of a fundamental time-direction degree of freedom has become the vital lead that allowed me to better understand many aspects of gravitational phenomena at the semi-classical level of description. Yet despite the fact that the main objective of this report was mainly to provide a consistent account of the way by which the concept of negative energy that emerges from those considerations can be integrated into a classical theory of gravitation, I have also made use of those theoretical developments to provide solutions to various specific problems in fundamental theoretical physics and cosmology.

First of all, using the proposed description of negative energy matter particles as consisting of voids in the positive energy portion of the vacuum I was allowed to show that the negative vacuum energy states which are already predicted to occur by conventional quantum field theory would not give rise to catastrophic vacuum decay and the creation of states of ever more negative energy. This is one clear benefit of the approach favored here in the context where the existence of those negative vacuum energy states must be recognized as unavoidable, even from the viewpoint of a traditional inter-

pretation. The prediction of an absence of vacuum decay can be considered as one of the most significant result of the alternative approach to classical gravitation which was developed in this report.

An important outcome of my analysis of the issue of discrete symmetry operations, on the other hand, is that it becomes possible to regain the symmetry that would be lost as a consequence of the imbalance between the number of matter particles and that of antimatter particles which is attributable to a violation of the symmetry under reversal of the direction of time, thereby allowing to avoid the difficulty which would arise in the context where this asymmetry could be related to the direction of time singled out by the thermodynamic arrow of time. This was achieved by recognizing that the condition of continuity of the flow of time along the world-line of elementary particles imposes a similar, but compensating imbalance between what we may describe as negative action matter and antimatter. This outcome is highly desirable given that it allows to avoid the conclusion that there would exist an absolutely characterized lopsidedness of the universe with respect to the direction of time in the context where the redefined time reversal symmetry operation T does in effect involve a transformation of matter particles into antimatter particles. This approach, therefore, allows to satisfy the requirement that there can be no absolutely characterized direction of time in our universe, such as might have arisen from the observed asymmetry between matter and antimatter.

Another significant outcome of my revised formulation of the discrete symmetry operations is the derivation of an exact binary measure of entropy for the matter contained within the event horizon of an elementary black hole. This result is particularly noteworthy in that it actually matches the constraints set by the semi-classical theory of black hole thermodynamics. The possibility that is allowed in the context of the proposed interpretation of negative energy states to generalize the analogy between classical thermal equilibrium states and black holes through an application of the concept of negative temperature can then be considered to merely confirm the relevance of the concept of negative energy matter for gravitation theory. Those unexpected benefits come in addition to the solutions which were offered in the first chapter of this report to the more traditional problems usually associated with the concept of negative energy matter and which allow to demonstrate the viability of a theory based on such an alternative interpretation of negative energy states.

But while the most decisive result derived in this report will probably re-

main the elaboration of a quantitative framework which generalizes relativity theory in a way that increases its simplicity rather than adding in complexity over the already elegant gravitational field equations, the most concrete results are those which were obtained by applying the lessons learned while solving the problem of negative energy states to address several long standing issues in theoretical cosmology. I believe that what this shows is that a cosmological model based on a consistent theory of negative energy matter provides a fertile ground for understanding all sorts of astronomical phenomena in which the gravitational interaction plays a crucial role. This appears nowhere more clearly than in the case of the cosmological constant problem. Indeed, using the proposed formulation of the gravitational field equations I was able to show that the cosmological constant, conceived as an average residual value of vacuum energy density, can be expected to be as small as the value of the scale factor determined using the metric properties of space currently experienced by positive energy observers is similar to that which is determined using the metric properties of space experienced by negative energy observers.

What makes this possible is the fact that additional contributions to vacuum energy density, arising from those zero-point fluctuations which are directly experienced only by negative energy observers, must be taken into account, which allow the natural value of the cosmological constant to actually be zero rather than the very large number associated with the energy scale of quantum gravitational phenomena which is produced by more conventional estimates. It remains, however, that even in the context where energy must be assumed to be null for the universe as a whole, a non-zero energy of matter can be compensated by an opposite energy of the vacuum and therefore, in the end, it appears that the fact that the cosmological constant is not much larger than it is currently observed to be can only be explained as being a requirement of the weak anthropic principle. Thus, one amazing consequence of this approach to the problem of dark energy is that despite the fact that it relies on the existence of a previously ignored symmetry principle, it nevertheless allows one to understand why it is that the cosmological constant was not perfectly null initially. But it also predicts that the average density of vacuum energy must actually be growing with time due to the divergence which necessarily develops between the rate of expansion of space measured by positive energy observers and that measured by negative energy observers, which is reinforced by the growth of the cosmological constant itself. This is certainly a result that will have decisive consequences for our understanding

of the evolution of the universe.

Also of importance is the conclusion that there must arise additional gravitational attraction on visible positive energy matter from the presence of underdensities in the negative energy matter distribution. What makes this conclusion particularly significant is the fact that those forces can be expected to have altered the gravitational dynamics of large scale structures in the primordial universe in ways very similar to those currently attributed to ordinary cold dark matter. In fact, it appears that this additional gravitational attraction, as well as the gravitational repulsion which would arise from the presence of overdensities in the negative energy matter distribution, may help explain certain aspects of the process of structure formation and certain phenomena taking place on the largest scale which have until now resisted a conventional interpretation.

It is quite amazing in any case that the additional source of gravitational attraction which arises from the presence of negative energy matter underdensities is allowed to so adequately complement the contribution to gravitational instability which is provided by ordinary dark matter, once the missing mass effect, which is usually believed to arise solely from the presence of weakly interacting massive particles, is understood to actually be a consequence of the presence of local variations in the density of vacuum energy which are attributable to the fact that opposite energy observers experience different metric properties of space in the presence of matter inhomogeneities. The conclusion that a much smaller portion of the missing mass effect can be attributed to baryonic dark matter particles which differ from their visible counterpart as a consequence of the fact that they carry opposite charges in opposite directions of time, is also significant, especially since it allows to provide an additional contribution to the fundamental binary degrees of freedom that characterize the state of matter particles on the quantum gravitational scale, thereby allowing to explain the fact that what once appeared to constitute a fundamental unit of surface actually contains four Planck units of surface, each of which can now be associated with one discrete degree of freedom.

But perhaps even more remarkable was the discovery that the existence of negative energy matter allows to explain the flatness of the universe without requiring one to rely on the uncertain theory of inflation, when it is recognized that energy must be null for the universe as a whole in order to satisfy the condition of relational definition of physical attributes. In such a context it appears that the presence of negative energy matter particles

described as voids in the positive energy portion of the vacuum is observationally confirmed, given that it is required to balance the energy budget of matter, while allowing gravitational energy itself to be null independently for positive and negative energy observers, so that the rate of expansion can be set to its critical value in the initial Big Bang state. This constitutes a further proof that the alternative concept of negative energy matter which I proposed based on independent motives is fully justified, even from a purely empirical viewpoint.

I must admit that for a while I was not fully convinced that a solution as simple as that which I had derived (based on the hypothesis of the existence of negative energy matter) could alone solve such a complex and difficult problem as that of flatness. What I had realized of course was that if I was right then it probably meant that inflation theory could no longer be invoked to solve other aspects of the inflation problem either and this was difficult to believe given that inflation theory was the dominant paradigm for cosmology at the time when I obtained my first results. But I came to recognize that this is the only appropriate conclusion and that there actually exists a more natural solution to the flatness problem that merely requires one to acknowledge the reality of negative energy matter. Thus, even though I may have preferred arriving at a different conclusion, there is no longer any doubt in my mind that it is really the condition of null energy (imposed by constraint of relational definition of physical attributes) and the balancing effect of negative energy matter which allow to explain the flatness of space on the largest scale[14].

It is while I was trying to solve the mystery of the thermodynamic arrow of time, however, that I was led to derive the most surprising results regarding classical cosmology and to gain the essential insights which allowed me to solve virtually all remaining aspects of the inflation problem. First of all, I provided decisive arguments to the effect that temporal irreversibility is not a matter of viewpoint, because the growth of entropy can be characterized in an objective way due to the existence of the natural definition of coarse-graining

---

[14]This is still true even though at some point I had come to believe that it was the weak anthropic principle that was constraining space to be flat, because I had realized that in a zero energy universe there must be a relationship between the density of vacuum energy and the curvature of space on a global scale, but I did not immediately recognize that the presence of this energy would actually make the universe flat, so that the condition of null energy alone is indeed constraining enough all by itself to produce flatness, as I had originally envisaged.

that is provided by the macroscopic parameters associated with black hole event horizons. But I also explained that there exists a usually ignored measure of information concerning the microscopic state of the gravitational field and that its value diminishes when the density of matter decreases below the average cosmic matter density. It is this variation that allows the total amount of information in the universe to remain constant even in the context where the amount of missing information required to describe the microscopic state of the gravitational field must be assumed to rise when local gravitational fields grow stronger and the area of event horizons grows larger. As a result, it becomes possible to conclude that information is always conserved, even in the context where expansion itself contributes a growing amount of missing information by continuously creating new elementary units of space in the vacuum.

Based on the notion that the thermodynamic arrow of time is ultimately attributable to the smoothness of the initial distribution of matter energy at the Big Bang, I was then led to propose that it is the requirement that there must exist causal relationships between all the elementary particles which are present in the universe that explains the asymmetric character of the growth of gravitational entropy. Most people no longer hesitate to recognize that the physical properties of our universe are constrained to a very small subset of potentialities by the requirement that those properties must allow for the spontaneous development of a conscious observer. What I have proposed is that solving that oldest of all physics problems, the mystery of the origin of the arrow of time, requires taking into consideration the similarly obvious requirement that for the universe itself to exist as a consistent whole, a certain requirement must be met which can only be satisfied when the universe goes through a state of maximum density and minimum gravitational entropy at least once during its history, because it is only under such conditions that all of its components can actually become causally related to one another. Thus was solved that long lasting puzzle. I believe that this unexpected outcome illustrates better than anything else the fact that there are serious consequences for assuming without any good reason the validity of certain commonly held hypotheses, such as the absolute positivity of energy and the purely attractive nature of gravitational interactions, because it is as a consequence of not having being held by such a prejudice that I was allowed to solve the problem of the origin of the arrow of time.

Historical perspective

The significance of the developments introduced in this report can be better appreciated by describing the progress achieved from a historical perspective. If we start with general relativity I think that what the theory allowed us to understand is that all motion, including acceleration, is relative and that the state of motion of an object must be defined in relation to the state of motion of the rest of the matter in the universe. Thus, relativity theory embodied in its structure the requirement of a relational definition of physical properties. But it also failed to integrate the requirement of the relativity of the sign of energy. The common belief which existed since the creation of the general theory of relativity is that energy must be considered positive definite, because otherwise apparently insurmountable problems would arise. Now, what quantum field theory allowed us to understand is that negative energies are unavoidable for properly estimating the probability of all possible transitions involving particles and antiparticles. But the current interpretation of this theory also failed to accommodate the fact that no constraint exists that would justify assuming that those negative energies are only relevant for computational purposes and do not show up as properties of real matter particles distinct from ordinary particles and antiparticles when gravitation comes into play. What I have tried to achieve in this report is to generalize relativity theory to produce a fully relativistic theory compatible with the requirement that the sign of energy should also be a relative property. What motivated those developments was a better understanding of the relationship between the sign of energy of a particle and its direction of propagation in time which again arose from applying the requirement of relational description of physical properties. In such a context it appeared in effect that a concept of negative energy distinct from that which is usually assumed to be relevant in quantum field theory was not only allowed, but was required by a truly consistent classical theory of gravitation.

Once it had been shown that the difficulties usually associated with negative energy matter can be solved without rejecting the physical relevance of the whole concept of negative energy, there appeared to no longer be any rational motive for rejecting the possibility that negative energies can propagate forward in time and give rise to gravitational phenomena distinct from those involving exclusively positive energy matter. It thus became clearly inappropriate to attribute a preferred status to positive energy matter and this in turn meant that we are no longer justified to assume (as some authors

did) that even as it becomes integrated with general relativity the concept of gravitationally repulsive matter cannot involve negative energy, but must merely give rise to the notion of a variable metric devoid of any theoretical justification. Indeed, it has been clearly emphasized in this report that it is only when the concept of negative energy is well integrated to classical gravitation theory, by considering the equivalence between the presence of negative energy matter and an absence of positive energy from the vacuum that a consistent theory (for which all measures of energy are relative) emerges which agrees with all experimental and observational constraints. The original approach which was developed in the preceding chapters is thus unique in that it actually allows to account for the very existence of the phenomenon of inertia, despite the fact that both positive and negative mass matter must be present on the largest scale. It alone also enables the success of the standard model of cosmology at predicting the rate of expansion of positive energy matter to be reproduced in a bi-metric theory.

It must be clear that the concept of negative energy already existed before the developments I proposed in order to make it a consistent notion were introduced. But negative energy was always defined in an absolute or non-relational manner which, as I have shown, leads to serious inconsistencies, in particular because it would give rise to violations of the principle of inertia. Indeed, the idea that energy could be negative in an absolutely defined way and should therefore gravitationally repel all matter regardless of its energy sign, as if this repulsion was a distinctive property of negative energy matter itself, was here shown to give rise to undesirable effects, even aside from the plain logical inconsistency it would involve. The alternative interpretation of negative energy states which I proposed has allowed to avoid those problems while also making unnecessary the hypothesis that only positive energy matter can exist, because it explains why matter in such a state is unobservable from the viewpoint of observers made of positive energy matter. It has also become possible to predict that negative energy matter cannot be found in large concentrations in regions of the universe occupied by positive energy stars and galaxies as a consequence of the mutual gravitational repulsion which must exist between particles of opposite energy signs and because negative energy bodies gravitationally attract each other. Thus, it was actually explained why negative energy matter has remained mostly out of reach of astronomical observations, so that this property no longer needs to be merely postulated.

It is the fact that the notion of time-directionality remained so poorly

understood, even after the progress which was achieved in this area by the creation of quantum field theory, that explains that there was so much confusion over what constitutes an appropriate definition of the discrete symmetry operations (from the viewpoint of both clarity and consistency) when I began studying the subject. It is indeed the stubbornness to consider time from a traditional viewpoint where only one direction is allowed for this degree of freedom that explains that the time reversal operation was never appropriately described. This was allowed to occur despite the clues arising from the discovery of antimatter and the successful description of antiparticles as particles propagating backward in time. The commonsense feeling inherited from our experience of thermodynamic time is so strong that it is still commonly believed that antiparticles are merely identical particles which happen to have opposite charges rather than being the same particle propagating backward in time, as seems to be required from a mathematical viewpoint. This is what explains that time reversal was never considered to involve a reversal of charge as I have shown to actually be required. But once this was recognized the possibility opened up to explain other facts. It is in effect by using this insight that I was able to propose an explanation for the fact that a finite number of discrete degrees of freedom, which is proportional merely to the area of a black hole, allows to completely specify the microscopic state of the elementary particles which were captured by the gravitational field of such an object. In such a context it can no longer be argued that the notion of backward in time propagation is merely an expedient for facilitating the calculations of probability amplitudes. Our notion of time direction has been irretrievably altered and there is no going back.

The remote future

To conclude this discussion I would like to offer a brief outlook on the far future of our universe as it emerges from the developments introduced in this report. Concerning first vacuum energy, what can be expected to occur from the viewpoint of both positive and negative energy observers is that the positive cosmological constant will keep growing. But if the amount of matter in the universe is infinite then the growth of local inhomogeneities in the matter distribution and the formation of ever more massive black holes of positive and negative energy signs will also persist indefinitely. From the viewpoints of both positive and negative energy observers the cosmological constant will

become arbitrarily large, but for a negative energy observer that will only occur after space stops expanding and the average matter density actually begins to grow. Now, if it was not for the fact that inhomogeneities are growing along with vacuum energy we might expect that this contraction of space would give rise to a final Big Crunch. However, given that from both the viewpoint of a positive energy observer and that of a negative energy observer local inhomogeneities in the matter distribution will keep growing, while the positive and negative energy matter distributions will become ever more polarized along energy sign, then it seems that even from the viewpoint of a negative energy observer the volume of space may never reach minimal proportions. In fact, under the influence of the gravitational repulsion between opposite energy black holes, space may eventually stop contracting and perhaps even start expanding again from the viewpoint of a negative energy observer, which would reduce the magnitude of the cosmological constant experienced by all observers to a lower level.

Thus, the outcome of the universe's current evolution should be a state in which matter is still expanding, while its overall distribution has become completely polarized along energy sign. Under such conditions the annihilation of all opposite action particles back to nothing (and the end of time which may be associated with it) will become more and more unlikely (even independently from the actual densities of matter which are involved) due precisely to the growing polarization of the positive and negative energy matter distributions along energy signs, which will prevent local contact between opposite action particles. Thus, despite its apparent tragic significance, the prediction of a growing cosmological constant may not give rise to a situation which from a practical viewpoint would be as hopeless as one might expect, because if gravitational entropy continues to grow as I expect, it means that a certain change will still be occurring on the largest scale for even that portion of the history of the universe during which the vacuum will provide a dominant contribution to the energy budget and this may allow some kind of progress and evolution to persist that would make use of those irreversible changes in ways that may simply be too difficult for us to imagine at the present moment.

Acknowledgements

Bibliography

[1] P. A. M. Dirac. A theory of electrons and protons. *Proc. Roy. Soc. (London)*, A126:360, 1930.

[2] R. P. Feynman. The theory of positrons. *Physical Review*, 76:749, 1949.

[3] H. Epstein, V. Glaser, and A. Jaffe. Nonpositivity of the energy density in quantized field theories. *Nuovo Cimento*, 36:1016, 1965.

[4] H. B. G. Casimir. On the attraction between two perfectly conducting plates. *Proc. K. Ned. Akad. Wet. Ser. B*, 51:793, 1948.

[5] L. H. Ford and T. A. Roman. Averaged energy conditions and quantum inequalities. *Phys. Rev. D*, 51:4277, 1995. URL arXiv.org/abs/gr-qc/9410043.

[6] L. H. Ford and T. A. Roman. Restrictions on negative energy density in flat spacetime. *Phys. Rev. D*, 55:2082, 1997. URL arXiv.org/abs/gr-qc/9607003.

[7] M. J. Pfenning and L. H. Ford. Quantum inequalities on the energy density in static Robertson-Walker spacetimes. *Phys. Rev. D*, 55:4813, 1997. URL arXiv.org/abs/gr-qc/9608005.

[8] M. J. Pfenning and L. H. Ford. Scalar field quantum inequalities in static spacetimes. *Phys. Rev. D*, 57:3489, 1998. URL arXiv.org/abs/gr-qc/9710055.

[9] D. R. Brill and S. Deser. Variational methods and positive energy in general relativity. *Ann. Phys. (New York)*, 50:548, 1968.

[10] D. R. Brill and S. Deser. Positive definiteness of gravitational field energy. *Phys. Rev. Letters*, 20:75, 1968.

[11] S. Deser. Timelike character of gravitational field energy-momentum. *Nuovo Cimento*, 55B:593, 1968.

[12] D. Brill, S. Deser, and L. Faddeev. Sign of gravitational energy. *Phys. Lett.*, 26A:538, 1968.

[13] R. M. Schoen and S.-T. Yau. On the proof of the positive mass conjecture in general relativity. *Comm. Math. Phys. (Germany)*, 65:45, 1979.

[14] R. M. Schoen and S.-T. Yau. Proof of the positive-action conjecture in quantum relativity. *Phys. Rev. Lett.*, 42:547, 1979.

[15] R. M. Schoen and S.-T. Yau. Positivity of the total mass of a general space-time. *Phys. Rev. Lett.*, 43:1457, 1979.

[16] R. M. Schoen and S.-T. Yau. The energy and the linear momentum of space-times in general relativity. *Comm. Math. Phys.*, 79:47, 1981.

[17] R. M. Schoen and S.-T. Yau. Proof of the positive mass theorem. ii. *Comm. Math. Phys.*, 79:231, 1981.

[18] M. M. Nieto and T. Goldman. The arguments against 'antigravity' and the gravitational acceleration of antimatter. *Physics Reports*, 205: 5, 1991.

[19] H. Bondi. Negative mass in general relativity. *Reviews of Modern Physics*, 29:423, 1957.

[20] H. Reichenbach. *The Philosophy of Space & Time*, page 256. Dover Publications, 1957.

[21] P. J. E. Peebles. *Principles of Physical Cosmology*, pages 108, 296. Princeton University Press, 1993.

[22] T. Piran. On gravitational repulsion. *Gen. Rel. and Grav.*, 29:1363, 1997. URL `arXiv.org/abs/gr-qc/9706049`.

[23] G. D. Birkhoff. *Relativity and Modern Physics*, page 253. Harvard University Press, 1923.

[24] S. Hossenfelder. A bi-metric theory with exchange symmetry. *Phys. Rev. D*, 78:044015, 2008. URL `arXiv.org/abs/0807.2838`.

[25] J.-C. Lindner. Theoretical basis for a solution to the cosmic coincidence problem, 2006. URL `arXiv.org/abs/gr-qc/0603005`.

[26] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project. Measurements of omega and lambda from 42 high-redshift supernovae. *Astrophys. J.*, 517:565, 1999. URL `arXiv.org/abs/astro-ph/9812133`.

[27] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.*, 116:1009, 1998. URL `arXiv.org/abs/astro-ph/9805201`.

[28] W. G. Unruh. Notes on black-hole evaporation. *Phys. Rev. D*, 14:870, 1976.

[29] S. Weinberg. The cosmological constant problem. *Rev. Mod. Phys.*, 61: 1, 1989.

[30] M. Zumalacarrégui and U. Seljak. Limits on stellar-mass compact objects as dark matter from gravitational lensing of type Ia supernovae. *Phys. Rev. Lett.*, 121:141101, 2018. URL `arxiv.org/abs/1712.02240`.

[31] S. S. McGaugh, F. Lelli, and J. M. Schombert. Radial acceleration relation in rotationally suported galaxies. *Phys. Rev. Lett.*, 117:201101, 2016. URL `arxiv.org/abs/1609.05917`.

[32] A. Kashlinsky, F. Atrio-Barandela, D. Kocevski, and H. Ebeling. A measurement of large-scale peculiar velocities of clusters of galaxies: Results and cosmological implications. *Astrophys. J. Letters*, 686:L49, 2008. URL `arXiv.org/abs/0809.3734`.

[33] A. Kashlinsky, F. Atrio-Barandela, H. Ebeling, A. Edge, and D. Kocevski. A new measurement of the bulk flow of x-ray luminous clusters of galaxies. *Astrophys. J. Letters*, 712:L81, 2010. URL `arXiv.org/abs/0910.4958`.

[34] K. Land and J. Magueijo. The axis of evil. *Phys. Rev. Letters*, 95: 071301, 2005. URL `arXiv.org/abs/astro-ph/0502237`.

[35] K. Land and J. Magueijo. The axis of evil revisited. *Mon. Not. R. Astron. Soc.*, 378:153, 2007. URL `arXiv.org/abs/astro-ph/0611518`.

[36] S. K. Lamoreaux and J. R. Torgerson. Neutron moderation in the Oklo natural reactor and the time variation of alpha. *Phys. Rev. D*, 69:121701, 2004. URL `arXiv.org/abs/nucl-th/0309048`.

[37] J. K. Webb, J. A. King, M. T. Murphy, V. V. Flambaum, R. F. Carswell, and M. B. Bainbridge. Indications of a spatial variation of the fine structure constant. *Phys. Rev. Lett.*, 107:191101, 2011. URL `arXiv.org/abs/1008.3907`.

[38] E. Tryon. Is the universe a vacuum fluctuation? *Nature*, 246:396, 1973.

[39] R. Penrose. Singularities and time asymmetry. In *General relativity: An Einstein centenary survey*. Cambridge University Press, 1979.

[40] P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, et al. Planck 2015 results. XIII. Cosmological parameters. *Astronomy & Astrophysics*, 594(A13), 2016. URL `arXiv.org/abs/1502.01589`.

[41] C. R. Burns, E. Parent, M. M. Phillips, M. Stritzinger, K. Krisciunas, N. B. Suntzeff, E. Y. Hsiao, C. Contreras, J. Anais, L. Boldt, L. Busta, A. Campillay, S. Costellón, G. Folatelli, W. L. Freedman, C. González, M. Hamuy, P. Heoflich, W. Krzeminski, B. F. Madore, N. Morrell, S. E. Persson, M. Roth, F. Salgado, J. Serón, and S. Torres. The Carnegie Supernova Project: Absolute calibration and the Hubble constant. *Astrophysical Journal*, 869(1), 2018. URL `arXiv.org/abs/1809.06381`.

Index