A Philosophic Study of Non-conceptualized Auditory Sensations: Mental States as Functionally Interpreted Abstract Dispositions

Sevi Siljola
Department of Philosophy
University of Helsinki
Academic Dissertation
26.9.1999

ISBN 951-45-9142-9 (PDF version)
Helsingin yliopiston verkkojulkaisut
Helsinki 2000

A Philosophic Study of Non-conceptualized Auditory Sensations: Mental States as Functionally Interpretted Abstract Dispositions

Sevi Siljola, Department of Philosophy, University of Helsinki

ABSTRACT

The aim of the present thesis is to examine the idea that mental states and consciousness in general are nothing above and beyond neural processes in the human brain. To this end, prominent positions from the modern philosophy of the mind are critically reviewed. A detailed description of the neurophysiology and neuroanatomy of hearing is presented. Also a model for the auditory processing of non-conceptualized auditory sensations is outlined. Some fundamental notions of current philosois outlined. Some fundamental notions of current philosophical theories of the mind are compared with the neuroscientific facts presented.

The comparison indicates conclusively that most of the generally accepted and used conceptions of the modern pphe modern pp and used fundamental conceptions of the modern philosophy of mind are not in accord with current knowledge regarding the structure and functioning of the human brain. Therefore, a new theory as to what constitutes the mental is suggested. The theory presented considers mental states as functionally interpreted abstract dispositions, which lack independent ontological status. Mental states are seen as mere figures of speech used in our ordinary language, which thus exist only relative to other minds and conventions. The proposed dispositional account is found to be strongly supported by the set of neuroscientific data presented. It is argued that qualitative aspects of consciousness (subjectivity and qualia) are created by a special causal reualia) are created by a special causal re by a special causal relation (auto-connection) which the brain has with itself. It is suggested that in the case of the auditory system auto-connections are formed between the central executive and preconscious processes. It is also apparent that under some conditions, activations of the N1 and MMN generators can function as relatively reliable indicators regarding the occurrence of conscious states. N1 and MMN (mismatch negativity) are event-related brain potentials (ERPs) which are elicited by auditory stimuli (N1) and stiiiiimulus differences or changes (MMN).

Contents

Introduction 3

I The Concept of Physical Realization 5

1. The Psychoneural Identity Theory 7
   - 1.1 The Type/Token-Distinction 8
   - 1.2 The Topic-Neutrality Problem and the Criterion of the Mental 9
   - 1.3 Causal-Role Identity Theories 18
   - 1.4 The Property-Exemplification Account of Events 24
   - 1.5 Objections to Psychoneural Identification 27
2. Metaphysical Functionalism 34
   - 2.1 From Behaviorism to Functionalism 36
   - 2.2 The Multiple Realization Thesis: Brain States *vs*. Functional States 45
3. Anomalous Monism 49
   - 3.1 Davidson's Proof of the Irreducibility of the Mental 50
   - 3.2 Some Debatable Issues 56
4. Problems with Reduction 61
   - 4.1 Ernst Nagel's Inter-Theoretic Reduction 62
   - 4.2 On the Relation of the Mental and the Physical: From Emergence to Supervenience 65
   - 4.3 Obstacles to Mind-Body Reduction 80
5. Kim's View of Physical Realization 82
   - 5.1 Nonreductionist Physicalism as Emergence 82
   - 5.2 The Problem of "Downward Causation" and the Charge of Epiphenomenalism 85
   - 5.3 Type-Identical Reductions of Mental States 88
6. Futher Considerations 91

II The Neuroanatomy and Neurophysiology of Hearing in Humans 94

1. The Neural Coding of an External Sound Begins in the Inner Ear 96
2. The Brain in a Vat Example and the Limits of the Auditory System 96
3. Sound Waves Are Transduced into Electrical Signals by Hair Cells 99
4. Subcortical Pathways and the Multiple Realization Thesis 100
5. The Anatomy and Connections of the Auditory Cortex 104
6. The Auditory System is Structurally and Functionally Specified 104
7. Are Subsystems Only Passive Reactors? 106
8. The Origin of Cellular Properties 107
9. In Place of A Summary 109

III Auditory Information Processing 110

1. Mental Representations and the Demand for an Inner Representational System 111
2. Outlining the Main Problems with the Concept of Representation 113
3. Information Processing in terms of Cognitive Neuroscience 115
4. In Search of a Model for Auditory Information Processing 119
5. Auditory Sensory Memory: "Short" and "Long" Sensory Stores 122
6. Automatic and Controlled Information Processing in Audition 125
7. An ERP Component Called *Mismatch Negativity* (MMN) Reflects Change Detection in Auditio 127
8. Näätänen's Model of Attention and Automaticity in Auditory Processing 129
9. Concluding Remarks 132

IV What Is Essentially Wrong with Contemporary Theories of the Mind? 134

1. Brain States and The Constitution of the Mental 135
2. The Physical Realization of Mental States Extends over Time 137
3. Functional Characterizations and the Criteria of the Mental 139
4. Summary 140

V A Homunctionalist Interpretation of the Auditory System: Mental States as Abstract Dispositions 141

1. The Concept of Disposition 142
2. Homuncular Functionalism 146
3. Mental States as Functionally Intepreted Abstract Dispositions 147
4. The Problem with the Lilliputian Argument: Is Information Processing Conscious? 151
5. Explaining Subjectivity and Qualia 152

Conclusions 157

References 159

Introduction

During the years that I have studied philosophy it has become clear to me that philosophical practice *per se* has very little informative value with respect to the natural sciences. Philosophy has always been considered merely as a cultural study and not as a part of strict science. In my view, the philosopher's task has been to maintain a certain level of sophistication in a society by contemplating those abstract issues – such as metaphysics, ontology, ethics, and logic – which have no practical bearing on everyday life or science. Neither are these studies expected to result in anything, for philosophical problems are generally thought to be unresolvable. The only reason there are still professional philosophers is probably because it is a duty of a civilized society to be willing to show interest in and spend resources on such an intellectual futility as philosophy. The consequence of this attitude is that there have never been any genuine breakthroughs in philosophy. Philosophy has neither been appreciated nor respected by other scientists in the academic world.

Of course, the picture outlined above is much too gloomy to be taken seriously. In fact, even in milder terms it might be considered as an overstatement. Nevertheless, there might be some truth in it. My personal opinion is that philosophical practice in itself is nothing more than a mind-game for adults who seek intellectual challenges. However, this does not mean that the philosophic tradition is worthless or that nothing can be achieved by means of philosophical reflection. On the contrary, I have grown to understand that philosophy has a unique place among the natural sciences. For me philosophy has never been an actual member of the sciences in the sense that it could provide some new empirically verifiable or irrefutable truths. Neither do I believe that a highly-trained thinker can obtain genuinely new scientific knowledge by pure reasoning. Instead, I have always seen the value of philosophy to be that it operates on a meta-level in respect to the natural sciences.

What I mean by this is that the philosophic tradition can provide a scholar with an excellent set of tools for thinking in cases where there is some scientific problem which resists explanation. The lack of explanation can usually be due either to the fact that knowledge of the subject is incomplete or that all the relevant information is available but nobody knows what to make of it. In these cases, the introduction of philosophy can lead to surprisingly promising results or even resolve the problem for good. In other words, my belief is that philosophy appears to be fruitful only when it is combined with the natural sciences as a meta-practice, the purpose of which is to resolve some specific scientific problem. This is also the approach which is taken up in this study.

The mind-body problem, or (to express it in more modern terms) the relation between consciousness and the brain, is a perfect example of a subject matter that has persistently resisted all kinds of explanations. Philosophers of the mind have tried to tackle this problem for centuries mainly with conceptual tools. Before the latter half of this century the knowledge of the human brain has been seriously limited and filled with misconceptions. Therefore, the application of pure reasoning was pretty much the only viable way to approach the issue. During the last twenty years the situation has dramatically changed. The invention of highly-developed brain research methods (CAT scan, MRI, MEG etc.) has provided us with a more accurate and comprehensive knowledge of the structure and functioning of the human brain. This increase in knowledge has not gone unnoticed. The newly discovered possibilities of the neural sciences have inspired consciousness researchers to

join forces: neuroscientists, psychologists, linguists, as well as philosophers are nowdays participating actively in the debate. New interdisciplinary areas of research – such as cognitive science, cognitive neuroscience, and artificial intelligence research – have evolved as a consequence of this new trend. Nevertheless, the mind–body problem still remains.

One of the reasons for the failure to explain the mind–body relationship could be that all the fundamental issues, such as the question regarding the ontological status of consciousness, have traditionally been left for philosophers to contemplate. Philosophers of the mind have in turn been relatively uninterested in or even ignorant of newly discovered neural facts which might be introduced into to their studies. The purpose of this thesis is to reach across the huge divide between the philosophy of mind and the neural sciences. The general aim is to take some specific set of neuroscientific facts related to a certain aspect of conscious activity and compare it to relevant philosophical notions and conceptions. The goal is to update the basic views of current theories of the mind and to simultaneously complete neural theories with appropriate philosophical reflections if some parts or aspects of them lack explanatory power. My firm belief is that if this can be carried out successfully, the means by which the mind-body problem can be laid to rest will be in our hands.

This thesis is constructed according to the following outline. Firstly, a thorough description of the prominent views of the philosophy of mind is presented, with a description of their strong and weak points. It is argued that Jaegwon Kim's type-identically reductive theory of the constitution of mental states is the most preferable alternative. The whole of chapter I is dedicated to these reflections. In chapter II a detailed description of neuroanatomy and the neurophysiology of hearing is given. This description is further supplemented by chapter III, which provides a specific model for the auditory processing of non-conceptualized auditory sensations. Comparison is made between the presented set of neuroscientific facts and some of the basic notions and conceptions of the modern philosophy of mind in chapter IV. It is evident that the generally accepted philosophical views, which also dictate Kim's theory, are incompatible with present knowledge regarding the structure and functioning of the human brain.

Therefore, a dispositional account of the constitution of mental states is outlined in chapter V. The theory presented is found to be in perfect accordance with the presented set of neuroscientific facts, and the chapter demonstrates that non-conceptualized auditory sensations are nothing more than brain activity. It is argued that the neural science has found solutions to many traditional philosophical problems a long time ago. Philosophers of the mind have neither been capable of or interested in finding them. For instance, the proposed reductionist view is supported by the introduction of a genuinely new way of explaining the qualitative aspects of consciousness (i.e. *subjectivity* and *qualia*) in terms of purely neural occurrences. It is also shown that the criterion of the mental can be found within the brain itself.

Chapter I
The Concept of Physical Realization

In 1967 Hilary Putnam published an article entitled ”Psychological predicates”, which profoundly influenced the debate on the mind-body problem. In this paper Putnam expressed for the first time the idea of *the multiple realizability of mental properties*.[1] He stated that in addition to the human brain, mental states (such as pain) can be ”realized”, ”instantiated”, or ”implemented” by a variety of other neuro-biological structures and organisms. Since then, the concept has been broadened and sharpened. Results of artificial intelligence research have given strong indications that even inorganic systems, such as computers and other AI machines, might also be able to realize mental states. Furthermore, new neurophysiological studies have conclusively shown that the same mental states are not only realized by different brain processes in different persons, but that even a single brain has several independent realization bases for the same mental state. Putnam’s claim, nowadays commonly known as ”the Multiple Realization Thesis” or ”the Multiple Realization Principle” (=MR), has enjoyed wide acceptance among philosophers.

In formulating MR Putnam started a new era in the philosophy of the mind. First of all, he prepared the ground for functionalism, which is thought to be the most promising – and arguably dominant – position on the nature of mind. However, Putnam’s paper had yet another serious consequence. It was mainly due to MR that the psychoneural identity theory came to its end. This was especially true of so-called ”type-physicalism”, which had started to attract some serious attention in the late 1950s and early 1960s, but its life span turned out to be unexpectedly short. The reason was interpretations, which made MR the antireductionist argument *par excellence*. In fact, by the end of the 1970s most philosophers had abandoned reductive physicalism as a doctrine about all special sciences (including psychology) and hence ruled out all possibilities of mind-body reduction. The rise of antireductionism as the pre-eminent view of the mind-body relation was further supported by Donald Davidson’s ”Anomalous Monism”.

However, during the past ten years the climate has yet again been changing. Antireductionist views have in turn been subjected to severe criticism. One of the motives behind this change of heart has been a growing dissatisfaction with the philosophical concept of supervenience. This notion that was once enthusiastically welcomed and to which great expectations were attached now turns out to contain many unpredictable problems. The slow but inevitable decline of supervenience has had an ironic impact on the mind-body debate. Since the late 1980s, a number of papers have appeared investigating the possibility of reviving the old notion of reduction. It seems as though the discussion is about to take another u-turn.

One of those in favor of the resurrection of reductionism has been Jaegwon Kim. He began a well-planned attack on antireductionism in 1989 with an article entitled ”The myth of non-reductive materialism”. In this paper Kim presented the argument of ”downward causation”, which stated that all non-reductionist views on the mind-body relation are bound to end up in epiphenomenalism. Kim has later repeated and reinforced this notion several times.

---

[1] However, the term ”realization” appeared in Putnam’s texts earlier in ”Minds and Machines”(1960).

However, instead of offering mere counterarguments, Kim presented in *Philosophy of Mind* (1996) his own proposal for resolving the mind-body problem for good. Kim's solution is based on a variation of Putnam's Multiple Realization Thesis, which he has structurally restricted to the human brain and to other neurobiologically similar organisms. Kim formulates his own notion of the concept of physical realization, where mental states are type-identically reduced to physical states and processes.

The primary focus of this chapter is directed towards describing Kim's view of physical realization. However, I believe that a fair amount of groundwork is needed before the arguments Kim has presented will come into their own. That is why it is first necessary to take a look at the type-identity theory, or the psychophysical identity theory, which was set forth by Smart and a few others. I will also consider some of the principal objections most commonly directed towards this thesis. Secondly, I will present a more thorough notion of Putnam´s MR and outline the basic principles of the highly influential non-reductionist functionalism. Davidson's alternative antireductionist view, Anomalous Monism, will also be dealt with fully. Thirdly, I would like to stress that Kim dares to go against all the odds and propose a radically reductionist view. This provides reason enough to dwell for a moment or two on reductionism; I will briefly outline what reduction is all about, and what its major difficulties are. After all this, I will finally reconstruct Kim's notion of physical realization based on four principles.

It is quite obvious that Kim has accepted a huge and difficult challenge. By openly defending a type-identically reductive mind-body relation Kim is daringly exposing himself to attacks from a wide range of philosophers. First of all, he has to come up with credible answers concerning the general problems connected to the notion of mind-body reduction. Even if Kim succeeds in doing this, he will undoubtly receive criticism from anyone with views even slightly unfavorable to physicalism. As well as enduring the moral outrage of antireductionists, there is the more serious and seemingly insoluble problem – the Multiple Realization Thesis. In order to justify type-identical reductions, Kim has to somehow find a way to get around MR. Naturally, he is not going to abandon MR together with the whole program of functionalism in favor of reductionism – they are much too widely accepted and established. Kim's only choice is to install reduction into functionalism. However, in the philosophy of psychology functionalists are an exceptionally tight and homogenous group, a large majority of whom swear fealty to the autonomy of mental life. By bringing reductionism into functionalism Kim is taking a radical and courageous step away from the mainstream of contemporary philosophy of the mind. This means that he should have something substantial to show for it. At the end of this chapter I will evaluate how strong Kim's arguments really are, and what defects or gaps his reasoning might have.

1. The Psychoneural Identity Theory

”The psychoneural (or psychophysical) identity theory”, more commonly known as ”the mind-body identity theory”, became the most influential mind-body theory in the 1960s – at least for a short time. It was made famous by J.J.C. Smart’s essay ”Sensations and brain processes” (1959; see also 1978; 1994) and was further supported by the work of U.T. Place (1956) and H. Feigl (1958; see also 1971). This position advocates the *identification* of mental states and events with physical processes in the brain. The basic idea is that just as modern science has shown our ordinary experience of light to be only frequency-varied electromagnetic radiation, in the same way mental phenomena could simply be considered as neurophysiological (i.e. physicochemical) events.

The terms ”light” and ”electromagnetic radiation” naturally do have different meanings – just as ”Morning Star” and ”Evening Star” have in the Frege’s classic example – but they still refer to one and the same phenomenon. When this line of thought is applied to the case of mental states, one can argue that even though, let us say, ”pain” and ”C-fibre activation” have different meanings, they do pinpoint the same phenomenon. And this is exactly the central thesis of the identity theory: in reality our sensations of pain turn out to be only activations of C-fibres. (Place 1956, p.108; Feigl 1958, p.439; Smart 1959, p.163)

So, the psychoneural identity theory states that mental events are identical to processes in the brain. Our example of pain could thus be expressed in another way by using statements such as ”Pains are C-fibre activations” or ”Pains are C-fibre excitations”. However, these kinds of statements need substantial clarification. The first and probably the most significant specification concerns the notion of ”identity”. The term ”identical” can be used to describe a number of different relations. For instance, it can mean equality in magnitude (same weight), refer to some instances falling under a certain type or kind (copies of a book), or express identity between referents of two names, such as ”Socrates” and ”Xanthippe’s husband”. Ordinary and often used identities worth mentioning are also mathematical truths. They have the property of being known *a priori*: ”2 + 2 = 4” and ”7 = the smallest prime number greater than 5” are necessarily true in every possible world and hence empirical verification is needless. (see, e.g. Kim 1996, p.57)

Mind-body identities proposed by psychoneural identity theorists, however, are something entirely different. According to Place (1956, p.105), Smart (1959, p.165) and Feigl, but also to D.M. Armstrong (1968b, p.77) and D. Lewis (1980, p.123), mind-brain identities are *a posteriori* truths, which should be established by empirical neurophysiological research. This means that psychoneural identities cannot be formed solely on the conceptual basis by using the meanings of the expressions ”pains” and ”C-fibre excitations”; these concepts are independent, scientifically certified theoretical identities in the same way as ”water” and ”H_2O” are. Instead, in both cases the phenomenon in question (pain, water) is identified with its counterpart described in the theoretical language of science (H_2O, C-fibre excitation). This notion appeals to common sense; it is able to explain not only why man was able to know so much about water before he knew anything about its chemical composition, but also why pains have been a part of ordinary life long before the discoveries of the neural sciences.

It is important to emphasize that the psychoneural identification (if it is actually true) is thought to be *necessary* , or of ”strict identity”, as Smart calls it (1959, p.162). The necessity in question is governed by the

principle named ”the indiscernibility of identicals”, also known as ”Leibniz’s law”. It can be defined in the following way:

> If *X* is identical with *Y*, *X* and *Y* share all their properties in common – that is, for any property *P*, either both *X* and *Y* have *P* or both lack it (Kim 1996, p.58).

The content of this law is simple and straightforward. If *X* and *Y* are actually identical, there is no need to postulate the existence of two things instead of one. In practice, this means that if ”pain” and ”C-fibre excitation” share all their properties in common, they are one and the same thing – and this is exactly the position held by the psychoneural identity theorists. On the other hand, if *X* or *Y* has a property that the other one lacks, the identity falls apart (for more on the relation of Leibniz’s law and the identity theory, see Rey 1997, pp. 48–64). Thus many opponents of psychoneural identification have claimed that mental events contain properties which cannot be found from their physical correlates. The dispute over this issue still continues.

1.1 The Type/Token -Distinction

Now that the ambiguity of the term ”identity” has been removed, it is time to move on to other aspects of the psychoneural identity thesis that need clarification. Generally the identity theory can be understood in two ways, both of which have a unique position regarding the concept of ”event”. A view that is currently known as the ”token identity thesis” or ”token physicalism” has gained a notable amount of popularity during recent years. Token physicalism takes events to be basic particulars of the world, which may have properties and fall under a variety of kinds. According to this view, a feeling such as pain is an event that may be of two different kinds, mental and physical (i.e. ultimately the neural kind), but it can also have two kinds of properties, being a painful event and a C-fibre excitation. So, the controversial claim is that pain is both a mental and physical event, and that, in addition, these events are identical with each other. How is this possible?

The solution is based on the notion that a concrete pain event is an instance of two different event-types, the mental type and the physical type: every event that has a mental property has also some physical property. John Foster expresses the ”every mental event is a physical event” view in a very understandable form by writing that ”the event of someone’s pressing a switch may be the same as the event of his turning on the light; the event of someone’s moving a piece of wood from one square to another may be the same as the event of his checkmating his opponent” (1994, p.301). On the basis of this, token physicalists believe it to be correct to say that every pain event is identical with C-fibre excitation.

This sort of token identity thesis was made famous by Donald Davidson’s influential program of Anomalous Monism. One of the reasons behind the success of token physicalism is undoubtedly the fact that it can make a lot of sense of the causal relations between the physical and the mental – at least for those who insist on maintaining the autonomy of the mental. I will return to discuss Davidson’s view in detail later in this chapter (for a lengthier discussion of token physicalism, see Foster 1991).

The second interpretation of the identity theory is based on a notion of event that is commonly known as ”the type identity thesis” or ”type physicalism”. This position proposes the identity of mental types with physical types. The insight of this notion can be demonstrated more clearly by contrasting it with token physicalism. As mentioned earlier, according to the token identity thesis every mental ”event token” is identical to some physical ”event token”. In practice, this means that every mental event, such as a pain or an itch, is correlated with some brain process. Type physicalism, however, states something much stronger than this. Type identity theory claims that a certain mental ”event type”, such as pain, is identical to a specific physical ”event type”, such as C-fibre excitation. Correspondingly, another mental type, such as an itch, is in turn identifiable with some particular brain event or brain process other than C-fibre excitation. *So according to type physicalists, the statement ”Pain is C-fibre excitation” expresses a strict correlation between, say, pain in the left middle finger and a particular corresponding neurophysiological process, which can have no exceptions or variation*s. Token physicalists instead insist only that a pain event is identified with *some* brain event without setting any restrictions on what the neurophysiological process should be. (further discussions of type physicalism can be found in Shaffer 1991 and Hill 1991)

The type physicalist version of the identity theory is the classic formulation that was set forth and defended by Smart, Feigl, and others. It was mainly this type identity claim that caused the psychoneural identity theory to provoke so much indignation among antireductionists, and which finally left it outside the contemporary debate on the mind-body relation.

1.2 The Topic-Neutrality Problem and the Criterion of the Mental

The psychoneural identity theory has a feature that is of the utmost importance and which deserves to be addressed here. What I have in mind is a grave difficulty in the identity theory nowdays known as ”the topic-neutrality problem”(=TNP). This trouble was first brought to public awareness by Smart as ”Black’s objection” (1959, p.166, footnote 13), but it was not until D.M. Armstrong that TNP became subject to explicit consideration (see 1968b, pp. 76–79). To outline the matter, I wish to cite a passage from William G. Lycan, who has successfully encapsulated the essence of this problem:

> ...if an identity-statement such as ”My pain at *t* = the firing of my C-fibres at *t*” is substantive and *nontrivial* (as it certainly is), then **the two expressions flanking the identity sign must be associated with distinct characteristic sets of identifying properties**, in terms of which we make separate and dissimilar identifying references to what is claimed to be in fact one and the same thing. (Lycan 1987a, p.9; the bold type is added by this writer.)

So, we certainly know the identifying properties of the physicalist expression ”the firing of my C-fibres at *t* ”. But as Lycan points out, the real problem lies in finding the basis, or ”the sets of identifying properties”, for the mental side (”my pain at *t*”) of the statement. And where could one find a synonym for this expression, that would also address the identifying properties explicitly?

In fact, the debate over TNP is a part of a much larger discussion which consists of attempts to set some kind of criteria for the mental and the physical. Basically, this quest originates from a need to fulfil the fairly reasonable demand, that anyone who wants to carry out psychoneural identification must also come up with some conception of what is characteristically mental and what is physical. This point has been noted explicitly by Richard Brandt and Jaegwon Kim (1967, p.216), who have stressed that otherwise the terms ”mental” and ”physical” might be defined in such a way that the identity theory would end up in self-contradiction.

I

Probably the only philosopher who has not endorsed this difficulty is Paul Feyerabend. He has suggested that materialists should simply give up all statements that assert or imply the existence of the mind and use only physicalist language to describe the functioning of the brain (Feyerabend 1963; for a criticism of his position, see Mucciolo 1973). However, Feyerabend has had influential followers in the modern mind-body debate in the form of Paul M. Churchland (1981;1984;1996) and Patricia S. Churchland (1986) – who are the founders of the view known as ”eliminative materialism”. This theory proposes the ”elimination” of folk psychology on the basis that it gives a profoundly false description of the operations of consciousness. According to Churchlands, folk psychology should be replaced with a new – preferably computational – theory strictly based on the neural sciences (for a review of the neural-computational approach, see Churchland & Sejnowski 1989).

Other scholars have instead tried to come to grips with this puzzle and, indeed, a variety of different suggestions have been put forward as candidates for the final settlement. These attempts have revealed that there are two main stumbling-blocks in defining the criteria of the mental: (i) the claim asserting the logical independence of the concepts ”mental” and ”physical” and (ii) the heterogeneity of the mental domain.

The first of these problems is of a dual nature. First of all, the criteria for the mental should not be too *narrow*, for this would trivialize the whole distinction. Definitions that ensure the truth of any form of identification, such as ”the mental is what is infallibly knowable” or ”the mental is what lacks a spatial dimension”, are good examples of trivial conceptions. They would obviously define the physical to be whatever is fallibly knowable or whatever has spatial dimension, and by doing so, would foreclose the truth of any identification (Macdonald 1989, p.4).

Secondly, Donald Davidson (1970) and Mark Johnston (1985) have noted that the criteria should not be too *broad* either, for it would then include pretty much everything as mental. Examples of these kinds of criteria are, for instance, the definitions that make anything mental, if it is describable using mental terms or verbs (non-eliminably), or if there is a mental open sentence (=a sentence with a mental verb) true of that event alone. Verbs can in this case be counted as mental, if they express propositional attitudes such as believing, intending, desiring, hoping, knowing, perceiving, remembering and so on. Now, if we take everything that has a mental description to be actually mental, we end up in a position where our criteria cover much more than just mental events. Davidson has elaborated this problem by writing about an intuitively physical event, a collision of two stars in distant space:

> There must be a purely physical predicate ”*Px*” true of this collision...at the time it occured. This particular time, though, may be pinpointed as the same time that Jones notices that a pencil starts

> to roll across his desk. The distant stellar collision is thus *the* event *x* such that ”*Px*” and *x* is simultaneous with Jones’s noticing that the pencil starts to across his desk. The collision has now been picked out by a mental description and must be counted as a mental event. (1970, p.211)

In summary, it could be said that Davidson and Johnston are right in claiming that the ”broad” description of the mental does not really work as a criterion for the mental. Even at its best, the broad description can only state that a particular mental event is temporally related to every other event.

The other major difficulty in defining the criteria of the mental, the heterogeneity of the mental domain, is due to the fact that mental events seem to be separable into two quite different categories. Many philosophers of the mind are inclined to postulate a loose distinction between ”sensations” and ”propositional attitudes”. The first category is thought to consist of conscious experiences of the outside world and of one’s own body (e.g. pains and auditory or olfactory sensations), which have a qualitative, ”felt” content. However, also sensations which lack the felt aspect, but are still qualitative, are included in the first category. An example of these ”unfelt” mental phenomena are perceptual sensations, such as seeming to see something red. The second category is in turn thought to consist of mental states that are actually psychological relations between a particular person and a propositional content. These relations can be expressed by sentences like ”The Pope John Paul II believes the earth was created by God”, which contain verbs of propositional attitudes.(see Macdonald 1989, p.5)

Some philosophers, Charles Landeman (1964) among them, have applied alternative terminology in establishing the same taxonomy. Landeman, for instance, makes a distinction between ”direct” and ”indirect” perceptions. ”Direct” is meant to describe knowledge gained by means of the senses, while ”indirect perception” consists of knowledge of particular things, facts, or events gained by means of the senses, together with previously acquired information and beliefs (ibid., p.308). So, this distinction is based on the fact that sensations are thought to be only subjectively apprehended. Propositional attitudes, instead, seem to be also accessible from the third person perspective; that is, they can be observed by another person. However, this approach does not appear to be too accurate. D.M. Armstrong (1961, p.22) has admitted that the distinction is not absolutely sharp, and that there are some intermediate cases, which do not fall clearly into either of the categories (Sanford 1984, p.60). Hence, Landeman’s conception is not a plausible one.

Another set of criteria has been proposed to maintain the sensation/propositional attitude -distinction. According to Davidson (1974), Colin McGinn (1978) and Kim (1985), propositional attitudes either are or are not consistent with the context of a person’s other beliefs, expectations, behaviour, and so on. To put it more simply, propositional attitudes are rationally or irrationally held. These characteristics cannot be found from ”direct” sensations, which obviously makes this rationality/irrationality -division quite a worthy mark of propositional attitudes.

II

I characterized above some of the basic problems in defining the criteria or the ”mark” of the mental. In spite of the large variety of papers written on this subject, I believe only a few prominent, or at least promising, ”schools” of thought have been shaped by the discussion. The major divider of opinions between these approaches has been the clear-cut distinction between sensations and propositional attitudes.

The sensation/propositional attitude distinction facilitates the categorization of mental phenomena. On the other hand, it simultaneously makes the establishment of unified criteria for the mental, that would be unrestrictedly applicable to all kinds of mental states, quite impossible. This inevitable dichotomy in the mental realm has basically divided scholars into two camps, both of which propose largely different and independent ideas as criteria for the mental. The first group advocates an approach that could be described as *intentional* (see Knowles 1981; Kraemer 1984), and which is inclined to consider the issue through propositional attitudes. The other group's point of departure could be fairly described as *epistemic* (see e.g. Levinson 1983), for their approach prefers to tackle the problem concentrating on "direct" sensations. However, I will follow Kim's suggestion (see 1971;1996) and add one more group to the list; that is, those philosophers who propose *nonspatiality* as the mark of the mental.

The Intentionality Criterion. The intentional approach obviously leads back to Franz Brentano and to his famous work *Psychology From an Empirical Standpoint* (1874). In this book Brentano first introduced intentionality as the distinguishing feature, which ultimately separates mentality from the material world. According to him, mental phenomena have the exclusive property of "intentional inexistence" or of "including an object intentionally within themselves". In simpler terms, Brentano claims that a) *only conscious, mental states and events are capable of being directed upon objects of the outside world*, and that b) *only mental phenomena can refer to contents such as other mental states or images of physical objects* (that might not be in sight at the moment of imaging). Since then, this notion has been adopted as one of the recognized and unchallenged presumptions concerning the nature of mental phenomena.

I mentioned that those in favor of the intentional approach have taken propositional attitudes as their starting point in establishing the mark of the mental. However, the connection between the criteria of intentionality and propositional attitudes has not yet been explained. To deal with this question, we have to add one more distinction to an already overwhelming list of classifications.

I noted earlier that, in the case of the topic-neutrality problem, the root of all evil is the claim that psychoneural identification requires logically independent criteria for physical and mental phenomena. However, supporters of the logical independence claim are inclined to advocate a version of the identity theory, that is generally known as the "ontological" (Mucciolo 1974, p.169) or "metaphysical" (Macdonald 1989, p.6) approach, and which is seen to deal directly with nonlinguistic entities such as events, states and properties. Naturally, this version of psychoneural identification has not escaped criticism altogether. In fact, opponents of the ontological view have suggested another interpretation of the identity thesis that they believe avoids the problem of fulfilling the logical independence demand. This other version has come to be known as the "linguistic" (see Chisholm 1964/65, p.264; Mucciolo, ibid.; Kraemer 1984, p.131; Macdonald, ibid.) approach, which is thought to characterize the mental and the physical in terms of linguistic units such as predicates, sentences and languages. (For a distinction between the "linguistic" and "ontological" thesis of physicalism, see Hempel 1969)

The application of the linguistic approach gives its advocates a certain advantage over the proposers of ontological interpretation in setting the criteria for the mental. This advantage has been described by Jaegwon Kim:

> ...in particular, by drawing a mental-physical distinction with respect to linguistic expressions, the materialist may formulate a linguistic version of the identity theory to the effect that whatever is ”described” by a mental expression is also describable by a physical expression, or that a ”complete description” of the world can be given in physical language. Since he is not saying that mental expressions are identical with physical expressions, there is no need for independent definitions...(1971, p.324)

So, basically the linguistic approach allows one to set aside the difficult task of giving independent definitions of ”mental expressions” and ”physical expressions”. This is due to the fact that with the linguistic approach the distinction between mental and physical is drawn with respect to linguistic expressions –”mental” and ”physical” can be distinguished and described inside the language irrespective of the fact that the language itself can be physical or mental.

Now the connection between the intentionality criterion and propositional attitudes can be seen. Those modern philosophers, who have followed Brentano in setting intentionality as the criterion for the mental, have preferred the linguistic approach. In other words, they have not endorsed Brentano's original idea that mental phenomena are characterized by the ”intentional inexistence of objects” but, instead, *intentionality has been seen as the logico-grammatical property of linguistic expressions or more precisely as the distinguishing feature of mental sentences*.

An explication of this idea can be found from the one of the most influential defenders of the intentionality criterion, Roederick M. Chisholm, who writes in his *Perceiving: A Philosophic Study* :

> Let us say (1) that we do not need to use intentional sentences when we describe non-psychological phenomena; we can express all of our beliefs about what is merely ”physical” in sentences which are not intentional. But (2) when we wish to describe perceiving, assuming, believing, knowing, wanting, hoping, and other such attitudes, then either (a) we must use sentences which are intentional or (b) we must use terms we do not need to use when we describe nonphysical phenomena.(1957, pp. 172–173)

The insight of this passage can be captured in a single idea that Kim (1971, p.326) calls the ”Brentano-Chisholm thesis”: mental phenomena (containing psychological attitudes) can be described only using *intentional sentences*, but physical phenomena in turn are only describable by nonintentional sentences. So, according to Chisholm, sentences that express mental phenomena are intentional, which means in practice that they have certain logical properties (e.g. they contain verbs of propositional attitudes) that cannot be found in sentences that describe physical phenomena.

However, Chisholm's argument, along with the whole intentionality criterion, has two fatal flaws. To begin with, it should be noted that Chisholm thinks that psychological (or ”intentional”, if you prefer) sentences have certain ”logical properties” that separate them from physical expressions. What these properties should be, no-one – not even Chisholm himself – seems to know. In addition, we established earlier that reference to propositional attitudes and hence to verbs of propositional attitudes is not by itself an adequate criterion; ”direct” sensations can sometimes be considered as working examples of mental phenomena that do not contain propositional attitudes. Because linguistic versions of the intentionality criterion fail to cover (at least most of)

the mental phenomena that belong to the category of sensations, they have generally been criticised for being too narrow a set of criteria for the mental.

Another difficulty with Chisholm's idea of separating "logical properties" is that it seems to rule out the possibility of psychoneural identification. By stating that intentional sentences have peculiar properties that physical sentences lack Chisholm seems to imply that physical events are something other than mental events. In order to avoid this dualistic conclusion, Chisholm is forced to add a strong restriction regarding event identity: if two sentences describe the same phenomena, they *have* to be logically equivalent (1957, p.173;1955/1956). However, this sort of logical equivalence is much too strong to be accepted, and – as Kim points out – it has drifted quite a long way from the identity theory of Smart and others, which asserts only contingent, empirically established identifications.

Chisholm's view had its adversaries even in its own time (see e.g. Heidelberg 1966; Lycan 1969). However, I have no desire to start an archeological excavation and dwell on these past objections. Instead, the intentionality criterion in general can be considered from a much more interesting angle. What I have in mind is the challenge issued by cognitive science in its efforts to build artificially intelligent machines. If we hold that the essence of the concept of intentionality is that it defines the *reference or aboutness* of our thoughts and beliefs, then we can easily imagine some sort of elementary computerized intelligence that is programmed to answer the typed question "What is the capital of the United States?" by showing on the screen the sentence "The capital of the USA is Washington D.C.". Now, "Washington D.C." obviously refers to an actual city and the sentence itself seems to express a belief that this city is the capital of the USA, which means that the computerized intelligence's answer fulfils the requirements of being an intentional state. But, in practice, the computer's answer is only the outcome of some causal chain of electronic states that it is ordered or programmed to realize in the presence of an appropriate initiating cause (i.e. the typing of the question "What is the capital of the United States?"). Hence the computer has no self-reflective ability to apprehend what is going on, and what the meanings of the sentences are. So, can the computer's answer really be characterized as intentional?

If the computer is not in an intentional state (and it definitely is not), then we have to somehow show that its state is lacking something essential which can only be found from human mental states. The idea that the reference between words and their denoted objects is not enough to characterize the intentionality of the mental was first noted by Ludwig Wittgenstein. He stressed that the referential relationship, for instance, between my certain *belief* ("Washington D.C. is the capital of the USA") and the particular *state of affair*s (Washington D.C. **is** the capital of the USA) has to have a feature that separates it from the simple references between words and their objects (Wittgenstein 1953, p.177).

In fact, this sort of distinguishing feature has been formulated by introducing the concept of *content intentionality* (in contrast to referential intentionality), which claims that mental states have the peculiar feature of containing representational contents and meanings (see e.g. Kim 1996, p.21). This notion is based on the widely accepted idea, that since the human brain cannot apprehend external auditory sensations such as air-pressure changes (i.e sound) or electromagnetic waves (i.e. light) directly, it transforms them into mental representations that refer to the state of affairs of the external world.

But can content intentionalism really draw the line between intentional and non-intentional states and hence set the criterion for mental and physical states? Frankly speaking, it cannot, because it does not give any means to verify whether a certain system actually has mental representations. Hence the computer's answer "Washington D.C. is the capital of the USA" seems to meet all the requirements of being in a content-intentional state: nothing indicates that it does not have representative contents.

John Searle has tried to tackle this problem by suggesting an additional criterion that facilitates the tracking of mental representations. Searle proposes a distinction between *genuine or intrinsic intentionality* (obtained by human mental states) and *derived intentionality*, which is attributed to systems that seem to be intentional, but nevertheless are not in their own right (see e.g. Searle 1983; 1992). Our case of computerized intelligence is a good example of derived intentionality. The computer's answer "Washington D.C. is the capital of the USA" is not actually intentional in any way. However, it can be seen to have derived intentionality in consequence of the fact that the computer was programmed by entities that have mental representations, and hence their own genuine intentionality has been "planted" or "transformed" into the computer's answer. In practice, this derivation of intentionality has been mediated by the programmed sentence "Washington D.C. is the capital of the USA", which contains the meanings and intentionality of the mental processes and language used by its programmers.

Even Searle's formulation, however, does not manage to dispel all the shadows surrounding the intentionality criterion. My opinion is that his distinction beween genuine and derived intentionality is driven by the urge to rule out the possibility that artificial (read: non-human) systems might be intentional. However, Searle leaves the most important question open; that is, what makes human mental states intentional? Before this issue is settled, we cannot make hasty judgements about the intentionality of inorganic systems. The problem of intentionality is definitely not a simple one to resolve. Humans have sensations that have mental representations but do not seem to refer to anything or have any meanings. These kinds of cases are, for instance, pains and meaningless sounds (=sounds without semantic content). These exceptions by themselves are enough to undermine Searles's claim that content intentionality could function as the criterion for the mental.

Those exceptions also seem to lead to a position, where the existence of mental representations functions as criterion for the mental. This is due to the fact that everyone agrees with the notion that, for instance, human auditory sensations are mental states. However, those auditory sensations I mentioned can be only referentially intentional (=they refer to air-pressure changes), which means that they do not have any meaningful content. But what then makes these sensations mental states? I believe the correct answer is that they are simply "heard". In other words the human brain has transformed the air-pressure changes into a mental representation. Evidently, the concept of "mental representation" now becomes the key issue in defining the criteria of the mental.

In the case of auditory sensations, it is only natural to assume that "to hear something" is synonymous with "to have a mental representation". If a person is deaf, some essential part of his auditory system is impaired and hence the auditory system is unable to produce mental representations. When this definition is reversed, it follows that *any system, irrespective of the fact whether it is human or non-human, or even a system of an inorganic nature, that can have mental representations can and should be counted as possible physical realizers of mental states*.

This conclusion suggests that the criteria for the mental might not be found by means of analysing mental concepts and their peculiar properties (such as being an intentional state). Instead, it implies that the criteria for the mental are bound up with a system's capacity to produce mental representations; in other words, the system's physical structure and properties determine, whether it can have mental states or not.

I will return to this intriguing issue in chapter III and chapter IV and try to show that "mental representations" can be understood as qualitative (i.e. "felt") and subjective experiences, which do not exist in their own right, and which are purely a systems' physical processes or computational states. But next I am going to describe the second major view in setting the criteria for the mental, which is the epistemic approach.

The Epistemological Criterion. Behind any sort of epistemic approach there is always the notion that mental states and physical processes are knowable in fundamentally different ways. For instance, one can hear sounds but not the neural processes that produce these auditory sensations. This simple example is thought to be an illustration of the fact that physical states are accessible only from an "external", third-person perspective, while mental states can be known from an "internal", first-person perspective. Those in favor of an epistemic criterion for the mental, hence, often refer to such conceptions as "direct knowledge", "privacy", and "self-intimacy", when describing a person's knowledge of their own mental states.

The concept of "direct knowledge" means that mental states are knowable (i.e. known to be psychological states) without inference or support from observation or any other kind of evidence. (The terms "observation" and "evidence" are generally thought refer to beliefs or previous observations, which might provide additional information on the subject and mediate the experiencing of the"directly known" subject and hence take away the directness and immediacy of the experience.) However, the definition of "direct" does not seem to be explicit enough, for are not the immediate sensations of the external world as direct and without reference to beliefs as knowledge of mental states? To avoid this ambiguity, it has been suggested that the notion of direct knowledge should be reinforced with the claim of "privacy" or "first-person privilege", according to which only a single subject has direct access – and this subject is the person himself in whom the mental state is occurring (see e.g. Heidelberg 1966; Kim 1971, p.337).

Although the idea that a person has a direct and privileged access to his own mental states sounds very appealing to common sense, it certainly does not seem a very strong and general criterion for the mental. This is the way the idea of "infallibility" or "incorrigibility" regarding the knowledge of one's own mental states is usually added to the notion. The infallibility claim states that a person cannot make mistakes in deciding whether or not he has a certain mental state (see e.g. Kim 1996, p.17).

This claim, indeed, seems to have a lot of truth in it. In support of it, we can repeat the argument I presented earlier in conjunction with the claim that the existence of mental representations could function as the criterion for the mental. For instance, if a person has a mental state elicited by auditory stimuli, he certainly cannot make any mistakes in noticing (i.e. hearing) it. The person either hears the sound or then does not. If the person does not hear the sound, then there is something wrong with his auditory system or with some other part of his brain that participates in the production of the mental state. In other words, the person cannot hear the sound because his brain has a defect and is not able to produce the mental representation.

Nevertheless, the epistemic approach suffers from a one major drawback which makes it incompatible with type physicalism. It could be argued that if a person can be directly aware of his mental states but not of his brain events, then mental events are not brain events (see Nagel 1965). The epistemic approach can also be criticized from another angle. One of the key issues in current consciousness research is the distinction between conscious and unconscious neural processes. Now, if the epistemic approach is assumed to be correct, it has no means of dealing with the aforementioned distinction. How does a brain process, which is identified with some mental state, and that associated with an unconscious brain process differ? If this fundamental question cannot be answered in any way, the claims about the mind-brain relation do not have a very firm foundation. And this holds irrespective of whether the claims are reductionist or antireductionist in nature.

In conclusion, I would like to state that the epistemic approach actually has a great deal more potential than any philosopher of the mind could ever have dreamed of. New neurophysiological studies have found a means to make the verification of occuring mental states (or at least auditory sensations) also possible from a third-person perspective. In chapter III I will introduce an automatic neurophysiological process located in the supratemporal plane of the auditory cortex, which controls the attention-swiching mechanism in the human brain, and which determines what auditory sensations enter consciousness. This process elicits a brain wave and can hence be detected by a number of different research methods.

The Non-spatiality Criterion. This approach is really not a solution to anything. The non-spatiality criterion has merely a historical importance and I am raising it mainly as a matter of curiosity (for a more thorough review of this issue, see Kim 1971, pp. 329–336).

The notion that the mental lacks spatiality originates from Descartes, who wrote that ”extension in length, breadth and depth, constitutes the nature of corporeal substance; and thought constitutes the nature of thinking substance”(p.liii). The latter part of Descartes’ definition is not necessarily entirely misguided. I cannot imagine that anyone could have anything against the notion that what is characteristic of the mental is that it is capable of thinking. Correspondingly, there is nothing wrong in claiming that material objects such as chairs and rocks lack the ability to think.

However, the inevitable consequence of Descartes’ definition; that is, that the mental does not have extension (i.e. that the mental is non-spatial) puts the view in strict contradiction with materialism. If the mental is thought to be non-spatial, the relation between mind and body can be explained only in two ways. The first option is straightforward substance dualism (which actually was Descartes’ position), where the human is thought to consist of two entities, the physical body and the immaterial mind. The other option is to postulate the existence of some sort of angel that has, in addition to the immaterial mind, a non-spatial body. Both of these versions are, at best, ridiculous, and they demonstrate the inadequacy of Descartes’ view.

In conclusion, the non-spatiality criterion has no place or use in the modern philosophy of mind. I honestly believe that the progress made in the field of the neural sciences can provide us with much better tools for finding and establishing the criterion of the mental than any philosophical jargon. I have already mentioned the promising results from studies done in the area of auditory sensations. I will later return to them and discuss

the possibility that the criterion of the mental could be found from the brain itself, not from conceptual analysis. It might just be that the searches for the criterion of the mental have been carried out in all the wrong places.

1.3 Causal-Role Identity Theories

There is still one formulation of the type-type identity theory that we not yet have discussed. This view, known as ”the causal-role identity theory”, is based on the early work of D.M. Armstrong, presented in *A Materialist Theory of the Mind* (1968b), and in the writings of David Lewis (1966;1970;1972). This theory attempts to explicate the nature of mentality and its relationship to the physical by applying the concept of ”causal role”. The notion of the mental is characterized by the ability of mental events to produce certain types of behaviour (e.g. in case of pains) or by their ability to be caused by certain types of physical stimuli (e.g. tissue damage). In simpler terms, mental states are perceived as *mediators* in input-output relations, where the inputs and outputs can just as well be other mental states as physical behaviour or sensory stimuli. The essential idea of the causal-role identity theory is that (a) these input-output relations are *causal*, and that (b) the mental states mediating the transactions from the received inputs (e.g. sensory stimuli) to the response outputs (e.g. emitted behaviour) have *causal roles* in these relations (Armstrong 1968b, p.82; Lewis 1966, p.100).

The similarity with functionalism is evident; it is the central thesis of functionalism that mental kinds, such as pains and itches, are distinguished from each other on the basis of the distinctive input-output relations found peculiar to each kind. Actually, the causal-role identity theory is generally thought to be just another version of functionalism – and on very reasonable grounds. The basic principles are the same and neither of them contain in themselves the claim of type-type identity. However, philosophers with physicalist tendencies typically consider the causal-role identity theory, or functionalism in general, in conjunction with some sort of type identity thesis. This also applies to the case of Armstrong and Lewis, who have afterwards combined the identity claim of the mental and the physical with the theory. These physicalist favored variations of the causal-role theory and functionalism have sometimes been categorized under similar titles, such as ”causal-theoretical functionalism” (Kim 1996, p.104) or ”the causal role/functional specification theory” (Macdonald 1989, p.40).

I

Although Armstrong and Lewis had somewhat significant differences in their views, and even though Armstrong later refined his own theory of the mind in *Consciousness and Causality,* written together with N. Malcolm (1984; see also Armstrong 1980;1984; Rosenthal 1984), it is evident that they both took seriously the difficulties in setting the criteria for the mental and the topic neutrality problem that came with it. I believe this foresight resulted from the fact that TNP had been in circulation (or at least had been noted) in debates over consciousness for quite some time after Smart’s first introduction of the problem. Hence Armstrong and Lewis were better prepared for objections from those who were especially troubled by ambiguities in the description of mental concepts.

In his early work Armstrong defends the ontological, or metaphysical, version of the identity thesis. Armstrong states this clearly by saying ”it must be possible to give logically independent explanations...of the meaning of the two words ´mind´ and ´brain´”(1968b, p.77). His belief is rooted in a notion he shared with Smart and Lewis: psychoneural identifications such as ”the mind is the brain” are not logically necessary truths, but *contingent* identities that should be established and verified *a posteriori* by empirical research. In other words, if the mind-brain identity is of the same kind as the identity between, say, a ”DNA molecule” and a ”gene”, then the precondition of any empirical verification is an explicit description of the concepts involved.

In the previous chapter I cited a number of reasons and supporting arguments for this claim. Perhaps I should add to the list just one more remark that has been noted at least by Smart (1959, p.166) and repeated by Steven White (1986). The basic idea is that, if we do not find descriptions or verbal explanations for meanings of mental terms without departing from a physicalist world view, we have to admit that there are irreducibly mental properties. Armstrong has put it another way by implying the following: ”...we are certainly not aware of the mental states *as* the states of the brain. What then are we aware of mental states as? Are we not aware of them as states of a quite peculiar, mental, sort?(1968b, p.78).” Naturally, the answer to this problem is to find topic-neutral expressions for mental terms. And, in addition, if physical terms are definable *a priori*, this should also be the case for the mental ones. But how is all this to be achieved?

Armstrong’s solution is twofold. Firstly, he fulfils the logical independence claim by employing solely conceptual, or logical, analysis of the mental terms. Secondly, Armstrong preserves the contingency of the identifications by introducing an empirical element – in practice, a scientific scrutiny – by which physical phenomena are located as certain brain processes. To elaborate both of these components and their relationship with each other, it may be best to take a few steps back and consider them in conjunction with Armstrong’s whole theory of the mind.

As I mentioned earlier, causal-role identity theories attempt to describe mental states by means of analysing their ”causal role” in a certain input-output relation. In fact, Armstrong’s logically independent, a priori analysis of mental concepts is heavily based on this notion. Armstrong has explicated the relationship between mental states and causal roles in the conceptual analysis in the following way:

> The concept of a mental state is primarily the concept of *a state of the person apt for bringing about a certain sort of behaviour*. Sacrificing all accuracy for brevity we can say that, although mind is not behaviour, it is the *cause* of behaviour. In the case of some mental states only they are also *states of the person apt for being brought about by a certain sort of stimulus*. --- in many cases, an account of mental states involves not only their causal relation to behavior, but their causal relation to other mental states. It may even be that an account of certain mental states will proceed solely in terms of the other mental states they are apt for bringing about. --- [It should also be clarified that] the ”bringing about” involved is the ”bringing about” of ordinary, efficient, causality. (1968b, pp. 82–83)

Some aspects of the quoted passage might need further clarification. Nonetheless, I believe two specifications are in place.

Firstly, the *dispositional* nature of mental states needs to be emphasized here. Even though mental states are defined in relation to each other, Armstrong (1968b, p.86), as well as Lewis (1966, p.104), believes mental states to be ”pure” dispositions. This is meant in the sense that mental states have the ability to bring about

certain types of behaviour regardless of whether or not they ever actually realize that disposition. I will return to dispositions in Chapter 4 in which I will also place Armstrong's view of dispositions under scrutiny.

Secondly, it is important to remember that Armstrong proposes *type physicalism*. However, this first "logical" or "conceptual" analysis of mental concepts does not take any position on which way (type-type or token-token identity) defined mental terms should be identified with physical processes. There are good reasons for this. The purpose of the logical analysis, in the first place, is to give a topic-neutral definition of mental terms, which means two things: i) it should not imply mental states to be purely physical states, and that (ii) it should not characterize mental concepts in a way that would show them to have irreducible, mental properties. Macdonald (1989, p.44) has pointed this out to be the crucial difference that separates causal-role accounts from behaviorism. Behaviorism is a straightforwardly reductive notion that identifies mental states with the causal roles they have in certain input-output relations. Causal-role theories in turn only use the input-output relation as a means to carry out the conceptual definition of the mental. Mental states are defined as states having a certain causal role in the "bringing about" of a certain behaviour or other mental states – they are not identical with their causal roles! With reference to dispositions, Colin McGinn (1980) and Lewis (1966, pp. 101–102) have made the useful distinction that in causal-role accounts mental states have dispositions but they do not consist of them. In conclusion, the first component takes no position on the relation between the mind and the body; it is the second, empirical component of Armstrong's theory that contains the presumption of the type-identity thesis.

Armstrong's second component is the thesis that identifies the mind with the brain. However, he has not constructed any strong or appealing arguments in support of the identity thesis. I find it rather surprising that Armstrong went to so much trouble to find a way to circumvent the topic-neutrality problem, but in case of psychoneural identification he appeals only to common-sensical argument. Armstrong is particularly fascinated by the identification of the gene with the DNA molecule as an allegory for the psychoneural identification of the mind and the brain. According to him, in the same way that the concept of the gene (=a factor in human or animal linked to the production of certain characteristics in that person or animal) has been found by empirical research to be identifiable by the DNA (=a substance found at the centre of cells: deoxyribo-nucleic acid), the concept of pain (= a state of a person apt for bringing about certain behaviour, such as groans and whines) can be identified with a certain brain process (= activation of C-fibres) pinpointed by neuroscientific studies. (Armstrong notes that these identities are "theoretical" in a sense that no one can observe in practice how the causal chain starting from the DNA molecule actually brings about a trait such as the colour of someone's eye.) As Armstrong himself puts it: "once it be granted that the concept of a mental state is the concept of the person apt for the production of certain sorts of behaviour, the identification of these states with psycho-chemical states of the brain is nearly as good a bet as the identification of the gene with the DNA molecule" (1968b, p.90).

Now we are in a position to summarise Armstrong's view in totality. When the first component (a logical analysis of the concept of mental states) and the second empirical component (scientific scrutiny revealing the underlying neurophysiological processes) are contingently identified with each other, we have in our hands Armstrong's causal-role identity theory. These three steps construct an argument that could be explicated in the following way (Lewis 1972, p.207; see also Macdonald 1989, p.45):

(1) Mental State *M* = The occupant of causal role *R* (by logical analysis of the concept *M*)

(2) Brain State *B* = The occupant of causal role *R* (by theoretical empirical hypothesis)

(3) Mental State *M* = Brain State *B* (by the transitivity of =).

The third clause contains the type-type identity thesis, which justifies psychoneural identifications of mental types (such as ”pain”) and physical types (such as ”C-fibre excitation”). However, this third clause also happens to be the weakest part of Armstrong’s argument. Putnam’s multiple realization thesis alone is enough to rip the argument apart. Armstrong has no case in claiming that C-fibre excitation, or the human brain in general, could be the sole ”realizers” or ”instantiaters” of the feeling of pain, and, in fact, soon after publishing *The Materialist Theory of the Mind* he was forced to aknowledge this. As Armstrong later in his ”Self-profile” admits:

> Looking back on *A Materialist Theory* in the light of subsequent work by others I think that my greatest omission was the failure to consider the question of *type-type* versus *token-token* identities between the mental and the physical. A few moments’ thought about the matter would have shown me that there were problems in identifying, say, the type pain with a certain physiological type. Even if every pain is purely physiological state, it is not at all obvious that that which plays the causal role of pain in different minds, different species etc., must always be the same sort of physiological event. (1984, p.32)

Armstrong’s causal-role account in the presented form is vulnerable to all the same objections as other type physicalist views, and it is certainly not an acceptable formulation in its early form (see e.g. Nagel 1970). This difficulty has also led Armstrong to make significant changes in his later views. One possible refinement that would make Armstrong’s account more plausible, would be to replace the type identity thesis of the third clause with token-token identity. This change, however, would remove Armstrong’s causal-role identity theory from the type physicalist program that was set forth by Smart and Place.

II

David Lewis can undoubtedly be counted on as a proposer of a causal-role identity theory. In fact, Lewis published his theory of the mind in ”An argument for the identity theory”, which was printed in 1966, two years before Armstrong’s *The Materialist Theory of Mind* was published for the first time. Actually, Armstrong was even troubled by the fact that Lewis’ view was very similar to his own notion, and that Lewis managed to get his article out before Armstrong’s important early work got published in 1968 (see Armstrong 1984, p.31). The outlines of Armstrong’s and Lewis’ theories are, indeed, amazingly in close to each other. This can clearly be seen in a passage quoted from Lewis:

> The definitive characteristic of any (sort of) experience as such is its causal role, its syndrome of most typical causes and effects. But we materialists believe that these causal roles which belong by analytic necessity to experiences belong in fact to certain physical states. Since those physical states possess the definitive characteristics of experience, they must be the experiences.(Lewis 1966, p.100)

In addition to the general formulation of the causal-role account, Lewis’s thinking was along precisely the same lines as Armstrong on two issues: mental concepts need topic-neutral analysis and mental terms should be

described in terms of the causal roles they have in certain input-output relations. However, Lewis went on to execute this program in quite a different way. Lewis applied a method that was first introduced by F.P. Ramsey and further developed by Rudolf Carnap, and which Lewis himself prefered to call ”an elimination of theoretical terms”. Even though Lewis' approach is quite technical and its proper demonstration definitely requires more space than is at my disposal here, I am still going to attempt a very simple and elementary level analysis of the insight offered by Lewis' work.

Lewis' account could be said to consist of two main phases: i) the elimination of mental terms and ii) the statement of psychoneural identity. The first phase is achieved by collecting all the truths of folk psychology into a single sentence, where mental terms appear as names, and which in totality forms a theory, *T* (1970, p.68; 1972, p.). The mental terms of this theory are defined in terms of their causal relation to other mental states and behavioral responses. Though, it should be noted that *T* contains terms other than mental as well; some of the statements might be, for instance, reporting observations. In fact, Lewis states that ”some other term” can have any epistemic origin or priority ”provided we have somehow come to understand it” (ibid.).

However, the most difficult challenge of the first phase is yet to be met; that is, the elimination of mental terms. Lewis' proposal is to postulate a new term-introducing theory T_r, which contains the same observational and other statements as folk psychology, but where the mental terms are replaced with T_r-terms (1972, p.210). Now we have in our hands a postulation of the form:

$T_r\langle t_1 \ldots t_n \rangle$

This postulation contains only observational terms and T_r-terms (whose referents are defined in terms of their causal relations to each other and to other entities named by observational and other terms) but not mental terms, for they are replaced with new T_r-terms. According to Lewis, the postulation of T_r is false if any of the T_r-terms is without denotation. In other words, the postulate implies that, for each T_r-term t_i , the sentence ”$(\exists x)(x = t_i)$”, which says that t_i names something (1970, p.80).

The referents of the observational terms were known even before the postulate T_r was established, but so far we have had no clue as to how to deal with the new T_r-terms. What are their referents and how are they to be interpreted and defined? Lewis' answer is that first the T_r-terms should be uniformly replaced by variables by which we obtain *a realization formula of* T_r: ”$T_r\{x_1 \ldots x_n\}$”. When we ”Ramseify” this sentence by existentially generalizing over each mental variable, we obtain another sentence called the *Ramsey sentence* of T_r, ”$\exists x_1 \ldots x_n T_r\{x_1 \ldots x_n\}$”, which states that any (or at least one) *n*-tuple of entities that satisfies this formula is the *realizer* of T_r. On the basis of this we can interpret that $T_r(t_1 \ldots t_n)$ claims that there exists a unique *n*-tuple of entities that satisfies the open sentence $T_r(x_1 \ldots x_n)$, and that the terms $t_1 \ldots t_n$ designate members of this particular *n*-tuple (1970, p.81; 1972, pp. 210–212; see also Macdonald 1989, p. 47).

These conclusions also give us tools to formulate definite descriptions of the T_r-terms. For instance, the first member of the postulation T_r –t_1– can be defined as the namer of the first component of the unique realization of T_r. This can be represented formally in the following way (see Lewis 1970, p.87):

$t_1 = \iota^2 y_1 \_y_2 \ldots y_n \forall x_1 \ldots x_n \, (T_r\{x_1 \ldots x_n \_\_ y_1 = x_1 \& \ldots \& y_n = x_n)$

This definition is based on the notion that t_1 is naming an entity (the first component of the unique realization of T_r,) which is followed by some n–1 entities and which comprises the particular n-tuple of entities that realize T_r.

This definition brings the first phase of Lewis' account of the causal-role identity theory to its end. The brilliance of Lewis' approach should not be underestimated. In consequence of this "elimination of mental terms", Lewis has managed to define mental terms without the danger of falling into the trap of circularity and losing the appreciation of topic-neutral language. However, the second and undeniably more problematic phase of Lewis' argument, the identity thesis itself, is still to be considered.

Lewis thinks along the same lines as Armstrong that mental terms should be construed on the basis of the causal relations they play or fulfil in certain input-output relations. These mental concepts (i.e. mental types), were thought to refer to the properties of organisms (i.e. physical types), and they were to be constructed by theoretical descriptions. In practice this meant, that psychoneural identities, such as "pain = C-fibre activation", were the same kind of contingent, "theoretical identities" (established by scientific scrutiny) as the identity between "water" and "H_2O".

I already mentioned in conjunction with Armstrong's view that this kind of type-type identity is vulnerable to objections based on the multiple realization thesis. However, Lewis was very aware of this problem. In "Mad pain and Martian pain" (1980) Lewis takes the possibility of variable realizability into explicit consideration. He tries to solve the problem by stressing that *a mad man or a Martian might feel pain in a different way to humans because their organisms have different descriptions for pain and hence a different physical state plays the causal-role of bringing about the pain state*. Humans, Martians and mad men all have a peculiar physical structure that realizes pains (in humans C-fibre excitation and in Martians, the inflation of fluid cavities for example), but in respect of their own group, species, or "population" the realization is of the same sort, "normal":

> I and Armstrong claim to give a schema that, if filled in, would characterize pain and other states a priori. If the causal facts are right, then also we characterize pain as a physical phenomenon. By allowing for exceptional members of a population, we associate pain only contingently with its causal role. Therefore we do not deny the possibility of mad pain, provided there is not too much of it. By allowing for variation from one population to another (actual or merely possible) we associate pain only contingently with its physical realization. Therefore we do not deny the possibility of Martian pain. If different ways of filling in the relativity to population may be said to yield different senses of the word "pain", then we plead ambiguity. (1980, p.128)

I have to say that Lewis has thought this issue through much more thoroughly than Armstrong. By admitting that different systems might have different descriptions (i.e. different causal roles) for bringing about pain states and that physical realizers of those causal roles might therefore vary, Lewis takes a huge step forward. However, with all due respect, his answer does not convince. Even if we accepted structural alterations in different species and the exception of a mad man, the case would not be any different. The prime problem with type physicalism is not that different species or organisms might have different realization structures for mental states, but that even the human brain has numerous independent realization bases capable of producing the same mental

[2] The symbol "ι" is a definite description–operator.

states. In conclusion, Lewis gives no explanation of how a certain causal role, that fits the framework of a single description about bringing about pain states, can have several, variant physical realizations. In this respect his account has no more credibility than Armstrong's.

1.4 The Property-Exemplification Account of Events

Another version of the identity statement between mental and physical events was originated by Jaegwon Kim in the 1970s and has generally come to be known as "the property-exemplification account of events" (Kim 1973:1974;1976). This theory deserves only a very short and sketchy presentation, for the empirical facts presented in chapter II and chapter III will show this view to be little more than amusing. However, I have good reason for mentioning this theory. First of all, the property-exemplification account of events can be seen to make a commitment to another of Kim's views, known as "nomological monism" (however the theory itself does not imply this), which states that each individual mental event is a physical event, with the consequence that mental properties are nomologically correlated with physical ones. Nomological monism is obviously the opposite of Donald Davidson's Anomalous Monism, which in turn states that any kind of nomological relations between the mental and the physical are impossible. So, nomological monism gives an important perspective on Davidson's work. Secondly, although Kim himself has not defended nomological monism for quite some time, he still sees the property-exemplification account of events as a viable approach (see introduction in 1993). I believe it is about time to disabuse Kim of these assumptions.

I

The novelty that Kim's property-exemplification account brings to the identity theory is that identity is not seen to hold between mental and physical *properties* but mental and physical *events*, i.e. between exemplifications of those properties. This shift in emphasis is meant to avoid the objection from phenomenal properties, according to which mental states, such as "stabbing pain", have qualitative, "felt" contents (or properties) that cannot be found from physical states and events. Instead of properties identity is viewed to hold between exemplifications of properties at a certain time in specific entities.

This notion requires that the concept of event is used in a very broad sense, and it is meant to cover not only states, conditions, and changes, but also the *having* of a certain property. This looseness, hence, gives an opportunity to use the term "event" when referring to states and conditions *of* and *in* substance instead of just concentrating on concrete changes in entities. Kim takes this advantage and defines events and states as "*exemplifications by substances of properties at a time*" (1976, p.34). More explicitly stated, every event is thought to consist of three components, which are (i) a substance or an entity (the "constitutive object" of the event), (ii) a property exemplified by or instantiated in that substance (the "constitutive property" or the "generic event"), and (iii) a time of instantiation of the property in question. This component-structure states simply that some substance has a certain property at a particular time, which according to Kim's notation is expressed by the

form [*x*, *P*, *t*]. Kim's theory needs, in addition, two basic principles to state the terms regarding the existence or occurrence of events and to the identification of them. These principles are (ibid., p.35):

> *Existence condition*: Event [*x*, *P*, *t*] exists just in the case that substance *x* has property *P* at time *t*.
> *Identity condition*: [*x*, *P*, *t*] = [*y*, *Q*, *t`*] just in the case where *x* = *y*, *P* = *Q*, and *t* = *t`*.

So, according to Kim two events, let us say A [*x*, *P*, *t*] and B [*y*, *Q, t'*], are identical if and only if x=y, P=Q, and t=t'.

II

However, I believe it has not yet become clear, how the property-exemplification account manages to avoid the objection from phenomenal properties, and how the identity between the mental and the physical is carried out in this theory. In answering the question of phenomenal properties, Kim relies on the view that was first introduced by Smart. Mental phenomena do not have any properties that could not be found in the physical world. This means in practice that there are no such things as "stabbing" pain or "slight" pain, only mental events such as "having-a-stabbing-pain" or "having-a-slight-pain" exist (Kim 1972, p.183).

The application of Smart's formulation gives Kim a chance to bury mental properties in the component structure of an event. This "fade out" of mental properties is carried out by a clever manouvre. According to Kim's notation, "the constitutive property" (e.g., "having-a-certain-feeling") is "expressed by any open sentence, or...as functions from possible worlds to sets of individuals"(1976, pp. 36–37). So, basically these sentences, or functions, attribute empirical properties to objects at certain times. There is nothing extraordinary about this conception. However, the key element is to understand that these attributes are not the events themselves but that – as Cynthia Macdonald has pointed out – they function as the basis for describing events: "The property whose possession by the relevant object at the relevant time is a property of that object is constitutive of the event which is (i.e. identical with) the exemplification of it by that object" (1989, p.102). In other words, *mental* properties do not exist *per se*, for they are only subparts of the very component structure that exemplify them. The exemplification itself does not have any mental attributes and hence the objection from phenomenal properties is overcome.

I am certain that this difficult and rather obscure notion needs to be made much simpler before it becomes fruitful. The basic idea is that identity between two events can be stated if either of these events has phenomenal properties. This is according to Leibniz's law, which requires that identical events share all their properties in common. Phenomenal properties cannot be found in the physical world and hence their attribution to mental events prevents event-identity. Kim's view ensures that events do not have any mental predicates. It is undeniable that a sentence like "John is in pain" attributes an empirical property to a certain object, and that the expression "in pain" refers to a phenomenal property. However, the sentence "John is in pain" is not the one that is being identified with anything. What is being identified, instead, is the property "John-is-having-pain" (whose structural components are a substance [=John], a constitutive property [=being in pain], and a certain time *t*) and the event that exemplifies it by that object (i.e. by John). Hence, the identity is justified because neither of the events has mental properties – only one of their components does.

III

Even though the property-exemplification theory manages to circumvent the objection from phenomenal properties, it needs something in order to support the identity thesis. This issue brings us back to nomological monism. In order to be justified, the identification of events needs to be backed up by two controversial claims implied by nomological monism. Firstly, the identification requires a metaphysical position on the concept of event that does not itself threaten the identification. Secondly, the identified events must be linked with psychophysical laws. Both of these claims brings nomological monism – and the property-exemplification account which is connected with it – into direct conflict with the views that propose the anomality of the mental.

Both preconditions of the identification are interrelated and hence the source of their trouble is the same. I mentioned in the earlier section that the criteria of the mental could be set on the basis of intentionality. According to this approach, the paradigmatic mental states should be considered as propositional attitudes, i.e. psychological states with propositional content. Hence intentionality is thought to be the essential feature that characterizes the mental realm. By stating that the mental domain has properties unique to it, this conception ruins the viability of event-identification. In addition, because both domains have peculiar and ”holistic” constitutive principles and they are so heterogenous in relation to each other, the intentional approach is thought to similarly preclude the possibility of psychophysical laws. This is precisely the view held by Donald Davidson (see 1970). Kim’s nomological monism is meant to rock the boat in both respects. The property-exemplification account of events postulates no characterizing mental properties for events, for these properties are only subcomponents, ”generic events”, of the whole construction of an event. When the obstacle of phenomenal properties is removed the chances of event-identification and also the viability of psychophysical laws are regained (Kim 1985, pp. 200–204; see also 1984b).

If we accept Kim’s notion that identified events are linked with each other by psychophysical laws, there remains the question as to what kind of laws we are talking about. Kim has suggested that the identified events, say a mental state *m* and a neural state *n*, could be *nomologically coextensive* . The nature of this connection is characterized in the following passage:

> Suppose further that there are neural states, n_1 and n_2, which are *nomologically coextensive* with m_1 and m_2 respectively; that is, we have laws affirming that as a matter of law, n_1 occurs to an organism at a time just in case m_1 occurs to it that time; similarly for n_2 and m_2. Now the neural states, n_1 and n_2, being theoretical states of physical theory, have *conditions of attribution*, that is, conditions under which their attribution to an organism is warranted...To say that [e.g.] C_1 is an attribution condition for n_1 must be more than to affirm a mere de facto coincidence of C_1 with n_1 (or with warranted attribution of n_1); it is to commit oneself to a statement with modal force...(Kim 1976, p.205; see also Brandt & Kim 1969).

The idea of nomic coexistence means simply that if a certain set of physical conditions hold, a specific mental state *necessarily* occurs. For instance, if a person suffers tissue damage, and he is in all respects normal, healthy, and alert, and we can further establish by neurophysiological methods that his C-fibres are activated, we must attribute a pain state to that organism; that is, it *necessarily* follows that the person is in pain.

Kim has explicated this ”necessity”, or ”modal force”, that arises from the presence of appropriate conditions using the following argument (ibid.):

(1) Necessarily, if C_1 obtains, n_1 occurs.

We also have the psychophysical law:

(2) Necessarily, m_1 occurs if and only if n_1 occurs,

whence:

(3) Necessarily, if C_1 obtains, m_1 occurs.

In the same way we have:

(4) Necessarily, if C_2 obtains, m_2 occurs,

where C_2 is an attribution condition of neural state n_2.

The problem with this definition is the postulate (2), which states the psychophysical law. This statement is also the essence of nomological monism, which has been challenged, or, more exactly, has been undermined by Davidson and his advocates. I will consider later Davidson's view and the reasons that have led to such wide-spread acceptance for the rejection of psychophysical laws.

In conclusion, it should be stated that Kim's property-exemplification account has other difficulties that are not necessarily related to psychophysical laws. For instance, Kim's theory avoids the objection from phenomenal properties because of its metaphysical position on events that allows it to bury mental properties only as subparts of a whole event. However, this conception of events as "structured complexes" has been challenged (notably by Davidson), and if the critics have any truth in their claims the problem of qualia might not be resolved after all. Another weakness in Kim's theory is that identity is assumed to hold between properties. In conjunction with nomological monism the theory hence seems to propose type-type identities, which in turn makes it subject to every counterargument that has ever been formulated against type physicalism. Probably the most problematic aspect of the theory, however, is the fact that the identity thesis also requires identity between times of exemplifications. In chapters II and III we will learn that this conception is simply not possible, for mental states are not realized at single moments, but as a result of many different processes during a longer period of time.

1.5 Objections to the Psychoneural Identity Theory

I have so far concentrated on outlining the general position of the psychoneural identity theory and describing the most important variations of this approach. All this is meant to give grounds for putting Jaegwon Kim's own type physicalist claims in a larger perspective and to offer a proper means of evaluating their validity. Along the way we have encountered a series of problems that have haunted type-type identity theories and which have ultimately led to their demise. I believe it is useful to end our discussion of the psychoneural identity theory by quickly summing up the gravest of these difficulties, for anyone who proposes a type physicalist or even slightly reductionist view of the mind-body relation must somehow resolve, or at least circumvent, these problems. However, I am not offering a detailed discussion of the subject. We will later have plenty of opportunities to return these issues and dwell on them more fully.

The Problem of Qualia. This issue is probably one of the oldest and most frequently cited to bones of contention in debates between reductionist and antireductionist philosophers. The origin of the problem is the fact that mental states seem to have qualitative, ”felt” contents that the physical world profoundly lacks. For instance, a person can feel either a stabbing or only a slight pain, which implies that such properties as ”stabbing” or ”slight” might be attributed to events of the mental domain. Since these properties are peculiar to mental states and they cannot be attributed to or found from physical events, it has been concluded that the mental cannot be reduced to the physical: mental states are not physical states.

In my opinion, the debate over this issue has not developed significantly since Smart (1958) offered his view that there are no such things as ”stabbing” or ”slight” pain *per se*. What exists, instead, are only neurophysiological processes that create the feelings of ”stabbing pain” and ”slight pain”. However, as a result of Thomas Nagel’s influential article ”What it is like to be a bat” (1974) there has been a shift of emphasis, where the qualia problem has been tightly linked with the difficulty of explaining the subjective aspect of experiences. In practice, the qualia objection has been transformed into a view, which holds that mental states are irreducible to physical states, because the subjectivity (i.e. the idea that mental states are experienced from a unique point of view) which characterises mental states can neither be found from the physical world nor be explained in purely physicalistic language.

My belief is that Smart had the right starting point in endeavouring to resolve the qualia problem. I also believe that references to the subjective aspect of experiences do not manage to undermine the fruitfulness of his approach. I will show in chapter V that Smart’s approach can be developed further to also include the origin of the subjective aspect of experiences. I am also going to introduce empirical data that conclusively show that the current knowledge of brain functions supports Smart’s notion, and that it can even pinpoint those processes that create the qualitative feelings and subjectivity of sensations.

The Problem of Multiple Realization. Putnam’s influential claim is directed against the narrow–minded notion that only the human brain is able and entitled to have mental states. It succeeded in widening the debate over consciousness in two respects. Firstly, the idea that other neurobiological or even inorganic structures might be capable of realizing mental states paved the way for artificial intelligence research and other more technical approaches to consciousness. Furthermore, the notion had a second aspect that forced philosophers to accept a fact that had been known among neuroscientists for some time: the human brain has several independent physical realization bases for the same mental states. Both of these aspects regarding the multiple realizability of mental states meant hard times for reductionistically oriented scholars. It was no longer credible to say that a certain mental state was identical to a certain physical process, for there were now several physical processes available to enter the identity relation. Hence many antireductionists thought they had won the battle over the autonomy of the mental unconditionally.

In reality, only the second aspect of Putnam’s MR is any threat to mind-body reduction. The idea that different species or qualitatively divergent systems have peculiar descriptions for the realization of mental states is in perfect accordance with the identity thesis. This is because one can restrict identification to a specific structure. If an advanced computer or Martian had mental states, it would not stop us from stating that a person’s certain

mental state is identical with a certain neurophysiological occurrence in his brain. Divergent physical realizers have peculiar and incompatible job descriptions. We saw earlier that Lewis (see 1980) anticipated this notion by stating that causal roles that bring about certain mental states might differ among ”populations” (i.e. among humans, Martians, and mad men), and we will soon see that Kim also has his own version of this ”structural-specific” (or ”species-specific”) restriction.

The second aspect of MR is undeniably a death blow for traditional identity theories. The empirically verified fact that different parts of the human brain can produce the same mental states excludes all possiblity of the classic mind-body reduction. The reason for this can be stated quite simply. If a certain mental state, M, has two alternative physical realizations, P_1 and P_2, and M is thought to be identifiable with its physical realization base, then M has to be identified with the disjunction of its physical realization bases, $P_1 \vee P_2$, because neither of the bases is solely *necessary* but merely *sufficient* for the physical realization of M. In practice, we are in a position where we have to postulate an identity statement between a mental predicate, M, and a disjunctive predicate, $P_1 \vee P_2$. This is precisely the source of the trouble, for is it justified to question the extent to which disjunctive predicates can be considered to be genuine properties?

I am inclined to agree with D.M. Armstrong (1978, pp. 20–23), Elliot Sober (1984, pp. 91–96) and Elizabeth Prior (1985, p.72) who have claimed that disjunctive predicates are not genuine properties. Two points can be made to support this notion. Firstly, if we consider a situation where there are two systems, A and B, of which A has only property, P_1 and B has a disjunctive property, $P_1 \vee P_2$, there is no difference in the causal properties of these two systems. If A suddenly obtains the disjunctive property $P_1 \vee P_2$, it will not have any new causal powers in reference to its original state, where it only had the property, P_1. Secondly, the ontological status of disjunctive properties is rather obscure. If one accepts the psychoneural identity between M and $P_1 \vee P_2$, then it simultaneously implies that all disjunctive properties should be counted as genuine properties. This would in turn lead to an over-general and hence undesirable ontological position.

This second aspect of MR sets a challenging criterion for a credible identity theory, and honestly speaking, I have not yet encountered a formulation that could explicitly meet this challenge. I do not believe that it can be achieved by means of traditional identity theories either. Instead, what is needed is a radically new and extraordinary approach. We will later discover wheher there is any such approach available.

The Problem of Locating Mental States. This is a problem that only a philosopher would consider with a straight face. The idea is that, since brain states have locations in space and mental ones do not, the psychoneural identity theory is false. I am sure that everyone can sense the presence of Descartes' non-spatiality criterion here and hence identify all the weak argumentation that goes with it. This objection can be easily refuted by repeating the psychoneural identity thesis: mental states do not have locations in space because they *are* the brain states that realize them. The same argument can also be expressed by using Kim's property-exemplification account, where mental events do not have locations in space because they *are* the events that exemplify them.

However, there is good reason for dealing with this location problem. It forces us to explain how it is possible to feel pain, for example, in a right thumb, if pain states, along with all the other mental states, are only occurrences of certain neural processes? This question reinforces the need to follow Smart's approach, where the

physical processes themselves create the qualitative aspects of sensations. It is an indisputable scientific fact that sensations are produced by and in end-organs. They in turn receive information about the body and the external world from limbs and other sense-organs in the form of electrical impulses. This seemingly self-evident truth, however, still does not erase the location problem from philosophical discussions of the mind-body relation. As I will later show, mental states are not realized by single neurological processes at a certain time. Instead, a number of processes contribute to information processing over a longer period of time. Hence mental states cannot be identified with any specific physical processes in the brain. Besides type physicalism, this difficulty haunts all forms of antireductionism, for it excludes the possibility of saying that a mental state has a certain physical realization base to which it is irreducible.

The location problem is a blatant example of the fact that philosophers of the mind have ignored the neural sciences for too long a time. The information that mental states cannot be identified or be seen to co-exist with single physical states, events, or processes has been around for quite some time. But it still makes all the philosophical theories of the mind look rather ridiculous. It seems as though these philosophers have not been doing their homework. Anyway, I will return to this issue later and present a thorough review of this subject in chapters II and III.

The Problem of Rigid Designators. This objection is voiced by Saul Kripke (1971;1972), and it is based on a distinction between rigid and non-rigid designators. According to Kripke, rigid designators are such that they designate the same objects in all possible worlds (i.e. they are necessary), while non-rigid designators are in turn merely accidental (i.e. they are contingent). There are also worlds were the designation might fail (1972, p.226). For example, ”heat” and ”molecular motion” or ”pain” and ”brain state B” are rigid, but ”Clark Kent” and ”the mild-mannered reporter” or ”the cause of heat sensations” and ”the cause of Brownian motion” are in turn non-rigid.

It is clear why ”heat” and ”molecular motion” are designated rigidly. If one ever came across a similar world where heat was something other than molecular motion, then a) either the word ”heat” would have a totally different meaning to the one it has in our world, or b) the other possible world would have different physical laws and structures and hence would not be ”similar” in any sense anymore. This is due to the fact that rigidly designated identities are also *metaphysically necessary*. This notion denies the possibility that either part of the identity relation can exist apart from the other. If heat is identical with molecular motion (i.e. heat *is* molecular motion), heat cannot be anything other than molecular motion or *vice versa*. If heat was something other than heat, it would also be something other than molecular motion. Correspondingly, the metaphysical necessity does not bind accidental designations. ”Clark Kent” and ”the mild-mannered reporter” do not apply as a rigid designation because it is easy to imagine a world that is in every respect similar to our world, but in which Clark Kent has experienced a rough childhood and in consequence has grown into a savage, beast-like criminal. (see Levin 1975, pp. 150–151)

The more interesting question is what makes ”pain” and ”brain state B” rigid but leaves ”the cause of the heat sensations” and ”the cause of Brownian motion” only as an accidental designation? In order to crack this

puzzle, one has to get to the core of Kripke's argument. Kripke gives two conditions for rigid designations (or for necessary identities, if one prefers):

> (1.) rigid designations are discovered and known *a posteriori*, i.e. through empirical investigation (1972, p.323), and
>
> (2.) the a posteriori discovery of necessary identities is established *contingently* (ibid., p.326).

To illustrate these conditions, let us take a common identity "water = H_2O" as an example and name it as a rigid designation "*R* = *R*'". Now, according to Kripke's approach, we are in an epistemic situation where we do not yet know anything else about the water but that it is "what falls from the sky when it is raining" and that it is "what has a freezing-point of 0°C". At this point we do not know whether the description is rigid. There might be another world where water's freezing-point matched but it still would not be the liquid substance falling from the sky in the case of a rain-shower. Perhaps some other water-like liquid that looked and tasted like water was the substance of rain-showers, which would mean that the descriptions "what falls from the sky when it is raining" and "what has a freezing-point of 0°C" picked out two different entities in that possible world. (see Macdonald 1989, p.30)

Because we do not know whether or not the designation is applicable in all possible worlds, we cannot postulate the rigid identity relation "*R* = *R*'". Instead, we have to depend on the first condition and fix the references of "*R*" (="water") and "*R'*" (="H_2O") with "*D*" (="what falls from the sky when it is raining") and "*D*'" (="what has a freezing-point of 0°C") and find out scientifically if the descriptions "what falls from the sky when it rains" and "what has a freezing-point of 0°C" are (non-rigidly) identical in this world. If they are, then we formulate the *a posteriori* and *contingent* relation "*D* = *D*'".

However, this formulation is still a long way from the rigid identity "water = H_2O". Now, the transition from this contingent relation into a rigidly designated identity is accomplished by satisfying the second condition; that is, the identity "*R* = *R* '" is formed by a contingent association from "*D* = *D'*". This procedure is realized in the following way, as described by Michael E. Levin:

> ...we let "*R*" rigidly designate the *D*, "*R'*" rigidly designate the *D'*, and we then discover that "$(\_x)D(x)$" has fixed the same reference for "*R*" that "$(\_x)D'(x)$ has for "*R*'". $(\_x)$ $D(x)$ need not have been $(\_x)D'(x)$, and that is what "*R* could have turned out not to be *R'*" means. (Levin 1975, p.151)

In practice, we let "water" designate rigidly "what falls from the sky when it is raining" and repeat it in case of "H_2O" and "what has a freezing-point of 0°C". We learned that it might not necessarily be water that was falling from the sky, but that it could also be some other water-like substance. However, now we have an empirical verification (obtained by scientific scrutiny) of the fact that in the case of rain it is actually water that is falling from the sky, and that it is H_2O that is the unique chemical compound that has a freezing-point of 0°C. Because we have established that "$(\_x)D(x) = (\_x)D'(x)$", Kripke believes we can infer, or make an "associated discovery", on the basis of it that "*R* = *R'*" and hence state that the identity "water = H_2O" is necessary and a posteriori. All this can be encapsulated in an explicit form that has generally become to known as "Kripke's Principle" (=KP):

KP: A necessary condition for "*R = R'*" to be known a posteriori is that there be a "*D*" and a "*D'*" such that ¬ *D*(*R*), ¬ *D'*(*R'*), ¬ ((_*x*)*D*(*x*) = (_*x*)*D'*(*x*)), and "*D*" fixed "*R*" and "*D*" fixed "*R'*". (Levin 1975, p.152)

Now that Kripke's rigid/non-rigid-distinction has been described, one may return to the real issue, the identity theory. What has any of Kripke's argumentation got to do with psychoneural identification? Well, Kripke's Principle itself functions as a connection to the identity theory. As I mentioned earlier, Kripke thinks that mind-brain identities, such as "pain" = "C-fibre exitation", are rigid designations. However, Kripke only uses this notion to attack the identity theory. He states that rigid psychoneural identifications fail to satisfy the conditions of KP and hence fail to be known a posteriori (1972, p.338). How should this be understood?

To catch Kripke's drift, let us take few steps back. As described above, the necessary requirement that KP sets for rigid designations is that they need to have a contingent associated discovery. This condition was expressed by the statement "¬¤((_*x*)*D*(x) = (_*x*)*D'*(*x*))". In case of mind-brain identities, this means that if, let us say, "pain" and "C-fibre excitation" were to be considered as a necessary identity, there would have to be an occasion in some possible world, where the identity would not hold. In practice, this would mean that either the pain was felt without the corresponding C-fibre excitation or the C-fibre excitation itself occured without the feeling of pain.

By now, probably everyone has realized where the problem lies. From whatever angle one looks at this situation, the identity theorists seem to be cornered. If one admits the possibility that a mental state might occur without a realizing physical event, it simultaneously means accepting the "Cartesian intuition" (see e.g. McGinn 1977) that mentality is something distinct and independent in relation to the physical world. This sort of dualist position would not only undermine the identity theory but it would also be totally unacceptable to any contemporary philosopher of mind – irrespective of the position one maintains on the mind-body relation. On the other hand, if one admits that C-fibre excitation could occur without the associated hurtful sensation, then the identity theory must be rejected too. This notion would in turn break Leibniz's law. In addition, this latter view is not feasible. If C-fibre excitation *is* pain, it is absurd to say that it could occur without the hurtful sensation: How can pain not be pain? Finally, if the identity theorist does not manage to fulfil KP's a posteriori requirement, psychoneural identities lose their status as rigid designations and they become only accidental. Even in this last case, mind-brain identities have to be abandoned. Non-rigid designations do not hold in every possible world, and hence there might be a world where a pain state might occur without C-fibre excitation or *vice versa*. And once more the identity theory must be forcefully rejected.

Kripke's argument is essentially of an analytic nature. By this it is simply meant, that Kripke has no interest in neurophysiological facts as to how the human brain produces mental states. His intention is directed elsewhere. Kripke's tendency is, more or less, to remind one that more care should be taken in formulating modal arguments in support of psychoneural identities – especially if the identities are thought to be necessary ones. Naturally, Kripke has had his share of critics too (see e.g. Feldman 1973;1974;1980). In fact, the opponents of his view have accumulated quite an impressive arsenal of counterarguments and objections during the past twenty years. My own approach in this thesis could be fairly described as empiristic or at least as neuroscientific. Hence,

instead of mere philosophical jargon, I will concentrate on the objection that is tightly linked with the functioning of the human brain.

Many scholars, at least Bradley (1964), Campbell (1970), Kirk (1974), Stich (1981), and Lycan (1987; see also 1974) among them, have found that Kripke's argument seems to end up in a regression. Kripke's argument has two points that support this notion. First of all, the a posteriori requirement of KP claims that if mind-brain identities were necessary and hence true there would have to be occasions where mental states and correlated physical processes were distinct. One would have to occur without the other. Secondly, Kripke's conclusion that psychoneural identities fail to be rigid suggests again that there exist possible worlds where mental states and physical realizers might occur distinctly without one another. In conclusion, Kripke strongly believes that mental and physical states are distinct.

This is the conception with which the problem of regression starts. If "pain" and "C-fibre excitation" were thought to be distinct, it would force one to accept a situation where C-fibre excitation failed to produce the associated hurtful sensation. Because "pain" and "C-fibre excitation" would in this case still be identical, one would have to make some kind of distinction between "pain" and the "sensation-of-pain". Hence "pain" and "C-fibre excitation" could be identical without the associated hurtful sensation because C-fibre excitation failed to produce the "sensation-of-pain". However, this notion is not an acceptable one. I believe Lycan pinpoints the problem by stating that if this were the case, "a sensation-of-pain distinct from a pain itself would also have to be a brain state" (Lycan 1987). Then we would in turn be in a position to repeat the a posteriori requirement of KP and to claim that there might be possible worlds where sensation-of-pain occurred without the corresponding brain state or *vice versa*. Then one would have to postulate a distinction between "sensation-of-pain" and "sensation-of-pain-of-sensation-of-pain" and the reasoning could go on *ad infinitum*.

Neurophysiological facts support Lycan's notion. In chapter III we will see that not all auditory stimuli reach consciousness. There is a special process that switches one's attention to certain auditory stimuli and hence makes the decision to transform sensory inflow into mental representations. Hence we can irrefutably conclude that Kripke's objection is of no threat to the identity theory and that it can be laid to rest in the graveyard of history.

The Problem of Mental Causation. The origin of this difficulty is quite simple. In our every-day experiences we seem to be in control of our thoughts and movements and we can make plans concerning the future. In other words, the essence and the prerequisite of being a human being is that one can decide what to think and when to think of it, or that a person can move and control his limbs. So, our common-sensical observations and reasonings of human behaviour support the following claim: *consciousness has causal powers through which it can affect other mental states as well as the physical world.*

However, our knowledge of the physical world contests this view. If mental states are identical to physical states and processes, then it seems as though mental life and consciousness in totality is only the outcome of a long series of deterministic and causalist physical processes. In this case, free will seems to be only an illusion. In addition, from a scientific point of view the idea that consciousness could somehow have an impact on physical processes is not an acceptable one. For a start, it would break the Law of Conservation of Energy among

other basic principles of physics. Secondly, where could this interaction between the mental and the physical take place? I believe Descartes' historical solution, the pineal gland, shows the desperation of the interaction approach.

No healthy-minded philosopher proposes straightforward dualism anymore. However, it is easy to sympathize with the reasons that have often led to dualistic positions. The idea that any theory of the mind should be able at least to explain how mental causation is possible or (if it is not actually possible) how the illusion of it is created, makes a great deal of sense. I shall also take this as my own standard in formulating a credible philosophy of the mind.

Mental causation is not a side-issue in the philosophy of mind. For instance, we will find later in this chapter that Jaegwon Kim has taken this problem as a starting point in constructing his own type-physicalist theory of the mind. Hence we will later have plenty of chances to penetrate the mysteries of mental causation.

2. Metaphysical Functionalism

The term ”functionalism” is known and used in several disparate areas of culture including architecture, design, literary theory, anthropology and mathematics – to name but a few. Due to its wide range of applications, a variety of meanings and connotations have been attached to the term. Unfortunately, these different senses rarely have anything in common, and the doctrines they represent usually do not share any unifying features. However, functionalism in the philosophy of mind differs from the general picture in one respect. Although there are a considerable number of variations (see, e.g. Rey 1997, pp. 184–201), the general features and principles of the different functionalisms in the philosophy of psychology are in exceptional conformity with each other. Variations are more consequences of fine-tuning than drastic conflicts between views.

Functionalism has obviously been rated as a highly plausible theory of the mind. Otherwise it would not have survived the exhaustive scrutiny it has been subjected to during the past few decades. On the other hand, without its solid and uniform framework, functionalism would not have been raised to a pre-eminent position above all the other mind-body theories. Nevertheless, I am willing to divide the seemingly homogenous program of functionalism in the philosophy of the mind with the following classification. It will not only help us to apprehend what kind of theory functionalism really is, but it will also facilitate a sharper analysis of the vast field of functionalist theories of the mind.

What I have in mind is the repetition of the division that was first presented by Ned Block in an article called ”What is functionalism?” (1980). In this excellent paper Block distinguished between three different meanings of the term ”functionalism”. These three influential functional approaches in the philosophy of the mind are: i) *functional analysis*, ii) *computation-representation functionalism*, and iii) *metaphysical functionalism* (see Block 1980, pp. 171–173).

In the first sense, functionalism is understood as a research strategy or a method of formulating psychological explanations. This approach has been forcefully defended and developed by Robert Cummins (1975;1983; see also Fodor 1968). In functional analysis a system is deconstructed into its component parts or subsystems, each having its specific task (=function) and a certain capacity to perform that task. The capacities

and functioning of the system are then explained by the functioning and the capacities of the subsystems, and by their method of integration. For instance, the ability of a certain production line to produce, say, cars can be explained by slicing the whole assembly line into many small components. Each of these components, such as the mechanism that does the coachwork, has a simple function (i.e. installing the coachwork into the chassis) and a capacity to perform it (i.e. an adequate supply of chasses and coachwork and an exact program for carrying out the actual installation). Hence the capacity of an assembly line to produce cars can be explained by noting that it has subcomponents which produce coachworks and chasses, and which enable it to install the manufactured parts of the car in the right order and in the right way (Cummins 1975, pp. 186–187).

Functional analysis can also be applied to biological organisms. For example, an entire organism can be divided into subsystems (or smaller organs), such as the digestive system or the blood-vascular system. The functioning of a single subsystem is then analysed as (and explained by) the capacity of a certain structure or organ. The capacities and functioning of the whole organism can hence be explained in terms of its subsystems (Cummins 1983, p.29).

The second sense of ”functionalism” in Block’s division, computation-representation functionalism, is closely linked with the increased influence of cognitive science and Artificial Intelligence research. This approach has two key assumptions. Firstly, psychological states are systematic mental representations of the world. Secondly, mental states and mental life in general are productions of simple mechanical computations that could in principle be compared to the primitive operations of a digital computer. Basically, mental states and complex psychological processes are thought to be explainable in terms of elementary computations which involve mental representations (Block 1980, p.171). It has also been suggested that since mental representations are systematic, there must be some kind of internal code, a language of thought, by which the mind represents things. The strongest and also the most radical advocator of this latter view has been Jerry Fodor (1975).

The third approach, metaphysical functionalism, is the most important element of Block’s division. While functional analysis and computation-representation functionalism concentrate on providing explanations of psychological phenomena, *metaphysical functionalism is, instead, meant as a theory of the nature of the mind*. It is precisely in this sense, as a view of the nature of the mind, that functionalism has gained wide acceptance (see Block 1978, p.268). It should also be noted that this is the only framework within which functionalism is understood in a uniform sense. If these boundaries are crossed, the consensus of opinion concerning the specific content of functionalism becomes fragmented.

However, from the point of view of my thesis the question of the exact nature of the mind is of a high priority. Hence the success of metaphysical functionalism as the thesis for the nature of the mind makes it a promising and acceptable starting point for our considerations. For instance, the general and simplified formulation of metaphysical functionalism itself takes no stand on ontological issues. If mental states are characterized in terms of their causal roles in certain input-output relations (e.g. that pain is caused by tissue damage which in turn has the tendency to cause whines and groans or the tendency to escape it) the pain itself can just as easily be identified with a certain physical state (Block 1980, p.172). Thus, the formulation leaves open the question of whether one should prefer an antireductionist or reductionist view on the mind-body relation.

Before proceeding to consider the actual subject a few points need to be made. I must emphasize that in the previous division Block himself does not simply talk about different meanings or aspects of ”functionalism”. Instead, Block believes that he is describing three versions of functionalism –of three different ”functionalisms” (1980, p.171). I am inclined to avoid Block’s original intention, for none of the categories of Block’s division exclude the others. Actually, the case is just the opposite. For an adequate and comprehensive theory of the mind there is the requirement that it is able to cover all aspects of consciousness. Any sort of statement concerning solely the nature of the mind will not offer explanations regarding intentionality, subjective experience, and qualia by itself. A plausible theory of mind also has to be in accordance with (or at least somehow applicable to) the neuroscientific view of brain functions and the operations of consciousness. Loosely outlined metaphysical functionalism is much too vague a notion to be compared with complex neural processes. I will later show in chapters III and IV that a complete picture of consciousness cannot be achieved without the use of functional analysis and computation-representation functionalism together with metaphysical functionalism. However, from now on, when the term ”functionalism” is used it refers to metaphysical functionalism as the general theory of the nature of the mind which takes no position on the ontological status of mental states – unless otherwise stipulated.

2.1 From Behaviorism to Functionalism

As we will soon find out, the functionalist program has its own independent roots deep in the history of philosophy, in ancient Greece. However, it also has a close relation to behaviorism. In fact, functionalism was in many respects thought to be a descendant of behaviorism – or at least its improved version. Unfortunately, no general agreement exists on what the exact relation between these two influential views of the nature of the mind should be. Nevertheless, a complete understanding of the functionalist program requires that the shift from behaviorism to functionalism is somehow taken into consideration. Therefore, before proceeding to the historical facets of functionalism, I am going to shed some light on the different behavioristic movements and their impacts on the formation of functionalism.

I

The beginning of the behavioristic movement is generally dated as occuring in the early twentieth century. The birth of behaviorism was the consequence of a debate over the nature of psychology as a scientific practice. William James had earlier set a direction for psychology in his classic piece, *The Principles of Psychology* (1890). According to James, psychology was to be considered as an independent field of science, but its focus was to be restricted solely to mental phenomena, such as feelings, desires, and their conditions. James had evidently adapted the Cartesian conception of consciousness, where mental states were profoundly private and subjective. Beliefs, thoughts, and pains were cognitively accessible only to the subject that was experiencing them, not to anyone else. Although James’ work reinforced the scientific status of psychology, it simultaneously drove

psychology further away from the other natural sciences: publicly observable phenomena and repeatable tests could not be fitted within the scope of psychology.

One of the founders of behaviorism, J.B. Watson (see 1925), and the early adherents of his view, such as A.A. Roback (1923) and A.P.Weiss (1925) contested James' conception. Initially Watson saw behaviorism as a theory mainly about the proper methods of scientific psychology. As Gilbert Ryle writes in *The Concept of Mind* (1949, p.327): "[behaviorism] held that the example of the other progressive sciences ought to be followed, as it had not previously been followed, by psychologists; their theories should be based upon repeatable and publicly checkable observations and experiments". In order to fulfil the scientific criteria of public observability and testability, private and subjective mental states and conscious acts had to be thrown out from the focus of scrutiny. The inner mental life was replaced with behaviour. This change in focus simultaneously altered the task of psychology, which was now redefined as the prediction and control of behavior.

Concentration on animal and human behavior was (and still is) the common denominator of all behavioristically oriented movements and theories. However, the concept of behavior also created divisions within the behavioristic program. Two issues have been the main objects of disagreement. First of all, the concept of "behavior" itself is in need of clarification. Psychologists and philosophers have not been able to agree on what the exact definition of the term should be. This has resulted in a number of different formulations as to what should be regarded as scientificly observable behavior. The second bone of contention has been the relationship between behavior and psychological theory. Should behavior be considered only as a constituent of mental states? Should psychological theories concentrate solely on observational behavior and exclude inner mental states from psychological explanations? Or should psychological theories also refer to mental states? Various answers have been proposed as solutions to these areas of obscurity. However, all the different views can be classified under two main headings: methodological and philosophical behaviorism. I will outline their general features below and consider some of their key weaknesses.

Methodological Behaviorism. The scientific view of psychology, also known as "methodological behaviorism", spread throughout academic circles all over the world and dominated psychological discussions (especially in North America) until the 1960s. This program took as its main objective J.B. Watson's original task of establishing psychology as a credible branch of the natural sciences by concentrating on the prediction and control of behavior. However, methodological behaviorism took the scientific criteria of objective testability and predictability very seriously. In consequence of stressing the intersubjective aspect of scientific scrutiny, behavior became the only object of psychological theories. The function of behavioral observations were in turn considered to be the refutation or the corroboration of formulated theories. Consciousness with its subjective and private aspects evidently fell outside the scope of methodological behaviorism and therefore of psychological explanation.

Methodological behaviorism has numerous versions of different strengths. "Different strengths" refers to the degree to which observable behavioral data is intepreted as consisting only of *physiological responses* (i.e. an increase in blood pressure, an increase in the pulse rate etc.) and *bodily motions* (i.e. the movements of one's arms, doing press-ups in the basketball court etc.). There is also the possibility that *mental events* (thinking, believing etc.) and *actions involving bodily motions* (going shopping greeting a friend) might be counted as

behavioral data (see Kim 1996). For instance, some less radical views consider verbal reports as behavior and hence allow references to internal, mental states of psychological subjects. However, these views are not without problems. If the subject utters, ”I have a pain in my left thumb”, it is not at all obvious that we can take the semantical content of this sentence to be credible behavioral data. The subject might not be able to make a distinction between pleasure and pain or even know English properly enough to express their feelings correctly. Hence many behaviorists accept only the utterance of the words, not their meaning, as behavioral data.

The most radical versions of methodological behaviorism, also known as ”black-box theories”, do not only deny references to a subject’s internal mental states, but, in addition, insist that psychological explanations should neither refer to physical or biological states of organisms. For example, B.F. Skinner has claimed in his *Science and Human Behavior* (1953, pp. 28–29), that psychological theories should set aside investigations of the internal structures and processes (even though they participated in bringing about a certain behavior) and concentrate solely on formulating correlational laws to observational input-output relations. According to Skinner, references to neurobiological states do not increase the knowledge or the explanatory power of psychological theories and are hence useless. A more fruitful approach is, instead, to take consciousness as a ”black box” and to focus on stimulus conditions and behavioral responses and their interrelationships.

Philosophical Behaviorism. As previously described, most forms of scientific behaviorism take a very rigid position on the concept of event. Hence inner mental and even neurophysiological states tend to fall outside the scope of psychological explanation. However, there is another version of behaviorism that sees the relationship between mental states and behavior to hold in a totally different way. While scientific behaviorism crudely excludes mentality from psychological explanations in order to make room for observable behavior, a view often characterized as ”philosophical behaviorism” takes behavior only as *constitutive* of mental states.

Instead of considering behavioral responses and mental events as members of two mutually exclusive domains, philosophical behaviorism sees them as two interrelated events. For instance, certain kinds of behavioral patterns can be seen to require intelligence, in other words the workings of a mind. On the other hand, without observable behavior the mind has a hard time proving its existence. Hence, it is very appealing to think that a mind expresses (or exemplifies) itself through observable, controlled behavior, and that the same behavior ”constitutes” mental states (i.e. the mind) which has co-ordinated the whole event (i.e. behavioral response pattern) from the beginning.

The creation of philosophical behaviorism is rooted in the early 1920s, when a group of philosophers, mathematicians and scientists started to hold sessions and reflect on certain problems in science and philosophy. After a while, the club began to describe itself as ”The Vienna Circle”. The name originated from the belief that they were reviving and continuing a late nineteenth century Viennese tradition, which was originated by physicists Ernst Mach and Ludwig Boltzmann. The most vital and publicly active – although rather short-lived – phase of the Vienna Circle started in 1929 with the publication of a manifesto entitled ”The Vienna Circle; Its Scientific Outlook” (*Wissenschaftliche Weltauffassung, Der Wiener Kreis*). In this manifesto the founding members of the Vienna Circle (of which M. Schlick, R. Carnap, P. Frank, O. Neurath, F. Waismann, H. Feigl, E. Zilsel and V. Kraft were the leading philosophical figures) gathered together the fruits of their discussions and

launched a powerful philosophical program called ”logical positivism”. (for a brief history of the logical positivist movement, see e.g. introduction in Ayer 1959; Smith 1987)

The characteristic method of logical positivists was *the logical analysis of the language of science*. The foundation of this approach was based on a conception of language which originated from Russell but was explicitly formulated by Ludwig Wittgenstein in *Tractatus Logico-Philosophicus* (1922). The conception takes there to be ”elementary sentences” whose characteristic feature is that they either correspond to absolutely simple facts or then do not. If an elementary sentence matches with the prevailing state of affairs, it is true and if not, it is not true. (The only exceptions to this rule are of course *tautologies* and *contradictions* , which make no claims upon the facts and are hence in agreement with every possible state of affairs.)

However, the idea of an elementary sentence – or ”protocol statement” (Neurath 1932/1933; Schlick 1934; Ayer 1936/1937) or ”primary sentence” (Carnap 1932), as it has also been called – was only the starting point for logical positivists. The more famous part of their approach was contained in a particular interpretation of Wittgenstein’s view. Logical positivists stated that not only the truth value of an elementary sentence but also its meaning was provided by acts of empirical verification. ”Empirical verification” in turn meant public observational reports. This doctrine, known as ”the verifiability criterion of the meaning”, was the most important building block in the logical positivists’ program. It also became their calling card after it obtained sloganist formulations, such as ”*the meaning of a statement is established by the conditions of its verification*” (Hempel 1935, p.17), and spread throughout academic circles all over the world.

The verifiability criterion of the meaning can be seen in action in Hempel’s following example:

> The theoretical content of a science is to be found in statements...Let us therefore ask what is it that determines the content of –one can equally well say the ”meaning”– of a statement. When, for example, do we know the meaning of the following statement: ”Today at one o’clock, the temperature of such and such a place in the physics laboratory was 23.4_ centigrade”? Clearly when, and only when, we know under what conditions we could call the statement true, and under what circumstances we could call it false...Thus we understand the meaning of the above statement since we know that it is true when a tube of a certain kind filled with mercury (in short, a thermometer with a centigrade scale), placed at the indicated time at the location in question, exhibits a coincidence between the level of the mercury and the mark of the scale numbered 23.4. (Hempel 1935, pp. 16–17)

To say that the temperature in the laboratory is such and such is the same thing as forming a proposition or a (protocol/elementary/primary) sentence, whose correspondence with the prevailing state of affairs can be tested. We can easily establish that the temperature and the whole situation expressed in the example-sentence occurs by placing a thermometer in the designated spot at the designated time. If the conditions that are expressed in the sentence hold and the thermometer shows exactly 23.4_C, then we can conclude that the statement is true. In addition to siply establishing the truth of the sentence, our observational verification also simultaneously constitutes the meaning of it.

However, it should be noted that the verification doctrine contains a few implicit and very important presuppositions. As was noted earlier, the essence of the natural sciences is that the phenomena they theoretically describe are publicly observable: the meanings of any scientific statement ought to be publicly verifiable through the act of observation. But what kind of phenomena then are publicly observable? Well, the obvious answer is

physical events. But do all scientific sentences contain only terms of physical phenomena which can be straightforwardly verified? In other words, are all scientific statements elementary sentences? No, they are not. That is why the conception of language that logical positivists adapted from Wittgenstein supposes that even though there are complex sentences, which do not behave like elementary sentences, they can be transformed or reduced to elementary sentences (Carnap 1932, p.63). And elementary sentences can always be translated into universal physical language (see Carnap 1932/1933).

For instance, in Hempel's example the word "temperature" can be totally eliminated from the the test sentence and be replaced by some *physicalist expression* such as "a tube of a certain kind filled with mercury, where the mercury level coincides with a certain number on the scale". As Hempel himself writes: "The statement itself clearly affirms nothing other than this: all these test sentences obtain...the statement, therefore, is nothing but an abbreviated formulation of all those test sentences" (1935, p.17).

By applying this method to the area of psychology in the early thirties logical positivists created a doctrine called "logical behaviorism" or "analytical behaviorism". The motives behind logical behaviorism were pretty much the same as those of J.B. Watson and other early behaviorists. The key issue was the unity of science, which meant in practice that psychology had to be established as an inseparable part of the natural sciences. The main goal of logical behaviorism was hence to bridge "the absolutely impassable gulf" (to use Hempel's expression) between the natural sciences and the sciences of mind and culture.

The means by which logical behaviorists tried to establish the unity of science were simple. Actually, there were only two basic principles. The first of these was that mental events were logical constructions out of behavior events. This idea was borrowed from Bertrand Russell (1924) who had suggested that one could get rid of some undesirable and philosophically troublesome entities, such as numbers, if they were considered to be only *logical constructions out of sets*. By introducing the notion that behavior is a constituent of mental states (i.e. mental events are *logical constructions* out of behavior events) they simultaneously laid the grounds for philosophical behaviorism in general.

The second principle was adapted from the general program of logical positivism. Carnap explicated this idea by stating that the unity of science would be obtained when "all empirical statements can be expressed in a single language, all states of affairs are of one kind and are known by the same method" (1934, p.32). In other words, psychological statements were to be treated no differently from other scientific statements. Psychological statements should be considered as ordinary scientific statements, which have to be verified by public observation. The verifiability criterion of meaning was hence to be applied also to psychological statements. As Hempel wrote:

> *All psychological statements which are meaningful, that is to say, which are in principle verifiable, are translatable into statements which do not involve psychological concepts, but only the concepts of physics. The statements of psychology are consequently physicalistic statements. Psychology is an integral part of physics*. (Hempel 1935, p. 18)

However, in the case of psychological statements, there is the conspicuous problem how psychological terms can be eliminated and replaced with expressions of physicalist language? The solution can yet again be found from behavior. The key is the observation of behavior and bodily processes. To be precise, behavior refers

to external bodily behavior ranging from plain hand movements to more complex speech movements. The bodily processes in turn include blood pressure, blushing, sweating, heart rate etc. Observations of these processes provide useful information about more subtle bodily processes and thus facilitate the verification of milder psychological reactions.

For example, by using these methods just mentioned the statement ”Jones has a pain in his left thumb” can be translated into physicalistic language in its entirety. Firstly, one can list some of Jones’ external bodily behavior: he is wincing, groaning and pressing an ice bag against his left thumb. One can also observe that Jones has a nasty cut on his left thumb, and that he keeps uttering the words ”I hurt my left thumb”. Furthermore, one can measure Jones’ blood pressure and notice a dramatic increase in it after he cuts his finger. Finally, one can monitor Jones’ brain functions using a PET-scanner (or fMRI) and find out that the brain areas, which are usually found to participate in the production of pain sensations, are activated. This list could be extended further, if needed. However, I believe the point has been made. The psychological expression ”pain” can be eliminated, for the meaning of the statement ”Jones has a pain in his left thumb” is nothing more than the fact that the described circumstances occur. In addition, if the verification is successful and the circumstances actually occur, one can also conclude that the state of affairs is such that Jones is in pain.

The general idea of logical behaviorism survived and became absorbed into other forms of philosophical behaviorism. This unifying feature was that mental states are defined on the basis of input-output relations or more precisely in terms of stimuli and response. Logical behaviorists thought that mentality was only a logical construct made from behavior events. In the same way other philosophical behaviorists did not place mental states somewhere in the middle of stimuli and response. For instance, pain states were not thought to causally bring about whinces and groans when sufficient cause of pain, such as cutting off one’s hand, occurred. Pains were neither attributed to particular states or occurences, for they were thought to be only descriptions of what one would do in case of certain stimuli. Mental states were, as one of the most respected philosophical behaviorist Gilbert Ryle described, ”pure dispositions”. By this Ryle meant that ”to possess a dispositional property is not to be in a particular state, or to undergo a particular change; it is to be bound or liable to be in a particular state, or to undergo a particular change, when a particular condition is realised” (Ryle 1949, p.43). This attitude denies that mental states have any part in input-output relations – they are only constructions out of those relations. We will soon find out that this conception became the major reason for the downfall of behaviorism and the rise of functionalism.

Troubles with Behaviorism. Interest in the behaviorist movement died down long ago and the sweeping influence it once had has now faded away with it. There are good reasons for this turn of events. A number of factors contributed to the failure of the behaviorist program. In case of methodological behaviorism, the belief that psychological explanations could be carried out solely on the basis of observable behavior without any reference to internal mental or neurobiological states turned out to be highly exaggerated. This idea, which was adapted from animal studies, was applicable to only the simplest cases of human behavior. A noteworthy explication of the unwarranted optimism of methodolgical behaviorism is Noam Chomsky’s justified and still timely critique of Skinner’s radical views concerning verbal behavior. Chomsky (1959) showed quite convincingly that learning

and the comprehension of language could not be explained without reference to the speaker's or the listener's internal mental (and neurobiological) states. The behavior of the speaker, listener, or learner of language is unavoidably linked with the conditions and contributions of the subject. Hence the radical versions of methodological behaviorism especially seem to be totally redundant in the analysis of more complex human behavior.

Philosophical behaviorism ended up wrestling with equally severe problems as did behavioral psychology. Logical behaviorists suffered from the same difficulties that troubled the general program of logical positivism. Probably the most insuperable problem relates to the doctrine of the verifiability criterion of meaning. The fact that verification acts themselves cannot be verified seriously undermines the credibility of their approach. Besides this specific methodological point there were other problematic issues common to all philosophical behaviorists. The pitfall was their fictional conception of mental states. Behavior events were thought to be constitutive of the mental which inversely meant that mental states were only (logical) constructions from behavioral input-output relations. Because mental states were only fictional constructions they naturally could not be causally efficacious or in any way part of the stimuli-response relations. Ryle's "pure" dispositions offer a striking example of this general attitude.

However, this notion was soon challenged. Roderick M. Chisholm (1957, pp. 181–185) and Peter Geach (1957, pp. 11–17) were among the first to point out that behavioral events themselves do not provide sufficient information for the definition of mental states in terms of input-output relations. The definition needs to be connected to other mental states – they have to *cohere* with the person's desires, beliefs, thoughts etc.

For instance, if Jones were famous for his craving for ice-cream, his desire could not be determined solely on the basis of observations, which reveal that Jones buys ice-cream every time he sees an ice-cream stand. A repeatedly occuring behavior in the presence of a specified stimulus does not suffice or justify drawing the following conclusion. Jones has to also *recognise* the stand as a stand (and not, e.g., as a pet store), *know* that that particular stand sells ice-cream (and not, e.g., soft drinks), and *believe* that the sales clerk will sell it to him, if he asks him to. It might very well be the case that Jones hates ice-cream but he believes it will lengthen his life in some mysterious way. Then the first conclusion that we draw from the stimuli-response patterns would be profoundly wrong. Jones does not crave, desire, or even like ice-cream. Instead, he consumes ice-cream on every possible occasion because of its believed beneficial, medical effects, which overcome the repulsion he feels for it.

Hilary Putnam approaches the subject from a different angle and notices the same defects in the behaviorists' argumentation. In an article titled "Brains and behavior" (1963) Putnam gives a number of detailed counterexamples which rather convincingly illustrate the weaknesses in behavioristic definitions of mental states. In his paper Putnam introduces "super-Spartans" whose culture and community encourages and requires that their members did not show any signs or make gestures that would indicate they were in pain when they actually were in pain. After this doctrine had been followed long enough and several generations had passed, super-Spartans had developed an amazing tolerance for pain. In addition, these highly developed super-Spartans had started to deny they had pains at all and supressed all talk about pains. Hence it is impossible to determine from the outside whether or not a member of this cult is in pain. But it is also impossible to construct any kind of pain state, even from verbal behavior, because their culture has erased all signs of pain. Putnam's example establishes that mental

states cannot be identified with certain dispositions to behave in certain ways in the presence of particular stimuli. Behavior events are not *necessary* conditions for pain. Conversely, we can imagine an analogous example where a person does not have the ability to feel pain for some reason, but he still pretends to feel it in cases where there are the appropriate stimuli. Hence he fakes his feeling of pain by wincing, groaning, and uttering the words which indicate he is in pain. This latter example in turn establishes that neither behavior events nor behavioral dispositions themselves provide *sufficient* grounds for the definition of mental states.

The philosophical behaviorists' most serious error was to exclude mentality from the bringing about of behavioral acts and other mental states. Today's action theories, instead, take it as given that explanations of human behavior require references to the subject's other beliefs, desires, feelings, etc. This still does not make behaviorism an embarrassing theory. By emphasizing the importance of stimuli-response relations in psychological explanations behaviorism paved the way for functionalism and the modern philosophy of the mind in general. When the role of mental states in input-output relations was taken as a target for re-evaluation, a major step of development towards the current view of the mind-body relation was soon taken. It became generally acknowledged that mental states were not constructions out of behavior events but that mental states were mediators in the transactions between inputs and outputs. When it was later observed that the mediating mental states were functional states instead of brain states, the first formulations of functionalism saw daylight.

II

The credit for formulating the very first version of functionalism has rightly been given to Aristotle. However, this interpretation ought to be accepted with certain reservations. The concept of consciousness was not used or even known until the 17th century. Hence the philosophers of antiquity and even of the Middle Ages were accustomed to speak of the soul (*psyche* in Greek; *anima* or *mens* in Latin). The term "soul" was not meant to refer to consciousness as we understand it but rather to the principle of life. As a product of his times, Aristotle too thought that a soul was something that made a thing a living thing. According to Aristotle's hierarchical view, even the simplest forms of life have "souls" or "psyches" of a lower degree. For instance, in the case of plants, the principle of life is their metabolism and their "psychic powers" are the nutritive faculties which keep the metabolism running. Man is naturally placed at the other end of the hierarchy. The characteristic principle of humans is our ability to reason and think: it is rationality that separates man from beast.

Now, the specific part which concerns functionalism can be abstracted from Aristotle's description. Souls or particular psychic states of living beings are not *entities*. Aristotle stresses that psychic states are only a means by which explanation is carried out. At every level of the hierarchy psyches are the modes of functioning peculiar to beings of each level. In the case of man, it might be said that the psychic states of the human being are the (functional) states which are used to explain his behavior. This interpretation makes the connection between Aristotle and the modern philosophy of the mind visible. Aristotle's description evidently bears a close resemblance to current formulations of functionalism. (Revonsuo et al. 1994, pp. 5–7; see also Hartman 1977)

Besides roots in antiquity, functionalism also has two historical strands that originate from the late 1950s and the early 1960s. These roots provide the link for the shift from behaviorism to functionalism. The first was set forth by J.J.C. Smart in his article "Sensations and brain processes" (1959). In this paper Smart opposed

behaviorists' tendency to strip away causal efficacy – and the status as real phenomena with it – from mental states. Smart insisted that mental states were actual phenomenal states that were produced by neurophysiological processes and which were also identifiable with them. Smart's new suggestion was that mental states were not abstractions from input-output relations, but that they occured in the presence of suitable stimuli and were caused by these stimuli.

Causal-role theorists, such as Armstrong and Lewis, later developed Smart's idea further and stated that mental states could be defined in terms of the causal roles they had in certain input-output relations. Mental states were hence not only caused by certain stimuli but they also participated in bringing about behavioral responses and other mental states: mental states were *internal* states of the brain. Armstrong and Lewis especially thought they were improving (and in some sense continuing) the behavioristic program. However, as was mentioned earlier, the general features of causal-role identity theory match quite perfectly with those of a loosely defined version of functionalism. So, those who see functionalism as a descendant of behaviorism are probably thinking about this line of development, in which functionalism is interpreted as an outcome of a transformation from behaviorism *via* causal-role identity theories (see e.g. Block 1978, p.268).

The other twentieth century source of functionalism has traditionally been traced to a series of articles written by Hilary Putnam during the early 1960s (1960;1963;1966). In these articles Putnam presented the first explicit contemporary formulation of functionalism, which is also considered to be the origin of the present-day debate over functionalism. Putnam's version was explicated and presented in terms of so called "Turing Machines", abstractly characterized computing machines created by the British mathematician-logician Alan M. Turing (for more on Turing Machines, see Davis 1958). However, nowadays functionalism is understood and described mainly by means of input-output relations, within which mental states play certain causal-functional roles. Since we have already dwelled a great deal on this latter issue, it is more convenient to set the issue of Turing Machines aside and concentrate on considering functionalism within the already existing frame of reference.

When functionalism is defined in a very general and non-controversial way, it states nothing more than that mental states are characterized in terms of their causal roles or causal relations to sensory stimuli, behavioral responses, and other mental states (see e.g. Block 1980, p.172). We will now explore this formulation more fully and take pain as our working example. There are certain requirements for an organism or for a person to feel pain. It might be said that the organism needs to causally respond to pain; it needs to have some kind of a damage-detector, which gets activated every time any sort of tissue damage occurs. This detector also has to have connections to other mechanisms that activate escape responses (such as the removal of a hand from a hot stove, when it is getting burned). When it is assumed that these circumstances pertain, one can try to define the conception of pain.

Let us think of a situation where a person accidentally spills boiling water on himself. The person naturally winces and groans, probably even screams loudly, and puts the burned area under cold water in order to relieve the pain and prevent even more severe damage to the skin. In this case, the concept of pain can be defined in terms of its *function* to a) detect the spilling of the water and the tissue damage caused by it, b) cause wincing, groaning, and screaming, and c) cause the placing of the damaged part under cold water. *Hence the concept of*

pain can be functionally expressed by stating that it is the causal intermediary between input (tissue damage) and the following outputs (wincing, groaning, screaming etc.).

However, this formulation of functionalism is in its vagueness not a very informative one. The first point that immediately draws attention is that the formulation does not indicate in what way functionalism differs from causal-role/functional specification theories. The deficiency of this definition can be revealed by a simple question: what is pain? The statement that the concept ”pain” can be defined in terms of its causal role in certain input-output relations does not by itself show what the concept is referring to. Are pain states immaterial mental states, brain states, or something else? This issue is of the utmost importance because the view that one chooses automatically decides which position is taken on the mind-body problem. Is functionalism a reductionist or an antireductionist theory? In order to answer this question, we have to once again introduce into this discussion Putnam's Multiple RealizationThesis.

2.2 The Multiple Realization Thesis: Brain States *vs*. Functional States

According to causal-role/functional specification theories mental states were to be characterized in terms of the causal roles they played in certain input-output relations. This is already an over familiar thesis, and it is well in accordance with almost any loosely defined version of functionalism. However, these theories contain a very powerful statement that distinguishes them from other functionalist theories of the mind: the statement of psychoneural identification. As I indicated earlier, functionalism (or metaphysical functionalism) is only a theory about the nature of the mind, which takes no stand on or preference to the debate over the reduction/antireduction issue. Causal-role/functional specification theories, in turn, take a clearly reductionist view of the mind-body problem. For instance, Armstrong and Lewis thought that mental states (e.g. as pains) were eventually identifiable with certain brain states (e.g. excitations of C-fibres).

Hilary Putnam's paper ”Psychological predicates” (1967) is considered to be a major turning-point in the modern philosophy of the mind. It gained its fame from the sharp critique directed towards the reductionist attachments of the functionalist theories of mind. In this paper Putnam introduced for the first time the Multiple Realization Thesis (=MR) into the debate over the mind-body problem. Putnam's argument was aimed primarily against the classic reductionist theories of mind, which presupposed that mental kinds (properties, event and state types) are correlated with physical kinds. A striking example of this attitude was Armstrong's and Lewis' psychoneural identity theory, which proposed the reduction of mental states to certain brain states. The argumentation by which Putnam went against classic reductionism is captured in the following, much-quoted passage:

> Consider what the brain-state theorist has to do to make good his claims. He has to specify a physical-chemical state such that any organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusc's brain (octopusses are mollusca, and certainly feel pain), etc. At the same time, it must not be a possible brain of any physically possible creature that cannot feel pain. (1967, p.228)

From today's point of view Putnam's argumentation contains only self-evident truths. The consequence of the correlation thesis, that a particular physical kind (i.e. a specific neural state) should be found to co-occuring with pain states in every possible pain-capable organism and structure, is undeniably rather ridiculous. However, Putnam was the first to recognize the contradiction with empirical facts to which the classic position evidently led. Although Putnam's approach was at that time more argumentative than based on strict empirical data, there are currently numerous research papers available from various scientific fields which corroborate his conclusions. For instance, I will show in chapter II that multiple realization is a natural part of human auditory information processing.

MR showed that the traditional reductionism presupposed an unacceptably anthropomorphic view of the causes of pain states and mental states in general. Once this error was revealed, the mind-body reduction seemed to lose its foundation. The basic problem was (and largely still is) that the reduction was thought to be based on an identity or a correlation between *properties*. MR makes it clear that there cannot be just one physical realizer (i.e. a property) *P* that would correlate with pain sensations in all pain-capable organisms. Instead, as Ernst LePore and Barry Loewer (1989) have noted, there are a number of different physical realizers $P_1, \ldots, P_n$ which are all peculiar to a certain species, and which are responsible for the pain sensations in these specific physical systems. In addition, we have established that there is neither a single neural correlate (i.e. a property) P_h which would solely cause pain sensations in humans. Instead, there is always a group of independent physical realizers $P_{h1},\ldots, P_{hn}$, which are all capable of causing pain sensations in humans on different occasions. Whether or not these sets are "open-ended" (i.e. whether they are infinite or finite) makes no difference on the metaphysical level.

The idea that pains could be identified with a certain neural substrate has inevitably come to an end. The reason for this is expressed quite convincingly by LePore and Loewer (1989, p.179), who have stressed that, since mental states have a disjunctive realization base, none of them is *necessary* for the production of mental states but only *sufficient* (see also Kim 1996, p.218). Hence the correlation between the mental and the physical does not hold between properties, but between a mental property *M* and a disjunction of two or several physical properties $P_1 v P_2 v \ldots P_n$.And this is the real root of the problem. I previously referred to the views of Armstrong (1978), Sober (1984), and Prior (1985), which all claim that disjunctive predicates are not genuine properties. If this is in fact so, the mind-body reduction between mental and physical states is clearly out of the question.

Frankly speaking, their claim has received a great deal of support. Besides LePore and Loewer, at least Hilary Putnam (1967, p.228), Geoffrey Hellman and Frank Thompson (1975, p.551), Richard Boyd (1980), John Post (1987), Ned Block (1990, p.146), and Derk Pereboom and Hilary Kornblith (1991) have been inclined to favour antireductionism. These aforementioned writers have adopted antireductionism for a common reason. The fact that mental states have several divergent physical realizers has three consequences:

a) since every member of any set of physical realizer properties is only sufficient for the production of mental states, none of them can be considered to be the base property into which the mental state in question can be reduced;

b) because disjunctive physical base properties are ”divergent”, neither can they be identified with each other;

c) hence the only way to carry out reduction from the mental to the physical is to identify a mental state with a disjunction of base properties.

As mentioned earlier, according to the traditional view, reductionism is seen as property identity. But if disjunctive predicates are not genuine properties, reductionism has to be abandoned.

But should disjunctive predicates be considered as genuine properties? No, they probably should not. In addition to the arguments I have already mentioned, disjunctive predicates have encountered a fair share of criticism from a variety of angles, mainly in favor of antireductionism. For instance, Jerry Fodor has claimed in his famous paper ”Special sciences, or the disunity of science as a working hypothesis” (1974) that there are no laws that could cover a reduction from mental states to heterogenous disjunction. Fodor’s critique is based on the notion that reduction should be carried out between mental kinds and physical kinds. Because disjunctive predicates are ”non-kind”, the reduction is bound to fail. Hence disjunctive predicates should be precluded. David Owens (1989, p.199) and William Seager (1991a, pp.93–98; 1991b, pp.126–130) are in agreement with Fodor stating that disjunctive generalizations are difficult to confirm. The difficulty arises from the fact that individual disjunctions are irrelevant to one another. For example, if a certain pain state is realized in a person by the neural state N_1 and N_2 (i.e. by a disjunctive predicate $N_1 \vee N_2$) and it is established that in another person the neural state N_1 also produces this same pain state, it still does not guarantee that the neural state N_2 will similarly produce this pain state in that other person.

I have presented here a number of arguments against disjunctive predicates. I must admit that I am inclined to believe they are correct: disjunctive predicates are not genuine properties. This conclusion, however, has serious consequences. As I have already made clear, if disjunctive predicates are not allowed, then psychoneural identification simultaneously gets ruled out. This leaves us with the following problem: If mental states are not brain states, what are they?

This question can be answered by first explicating what is meant by the identity ”the mental state M = the brain state B”. For a start, it should be noted that the fact of multiple realizability is not the only problem that haunts this sort of identity thesis. The troubles originate at an ontological level. The ontological picture behind psychoneural identification suggests that mental properties are ”first-order properties” in their own right. They have characteristic, intrinsic features, which physical properties lack. For instance, the nature of a mental property is to have a qualitative content, in other words to have a phenomenal feel. Furthermore, these mental states are thought to have nomological neural correlates in the form of brain states. However, MR refutes the possibility of a ”state-to-state” (or a ”property-to-property”) correlation. In addition, the view that presupposes an intrinsic nature to mental states is strongly leaning towards an unacceptably dualistic position on the mind-body relation.

Because of these reasons, Putnam and other functionalists propose a totally different sort of ontological view. In functionalist conceptions of mind the relation between the mental and the physical is not considered to be such that brain states ”produce”, ”cause”, ”bring about”, or ”correlate” with certain mental states. Instead, it is

thought that physical structures (e.g. the human brain) ”implement”, ”instantiate”, or ”realize” mental states. The choice of vocabulary adopted here is anything but an insignificant matter. Jaegwon Kim has quite comprehensively explained the requirement for a new set of terms by stating:

> There is the suggestion that when we look at concrete reality there is nothing over and beyond instantiations of physical properties and relations, and that the instantiation on a given occasion of an appropriate physical property in the right contextual (often causal) setting simply *counts as*, or *constitutes*, an instantiation of a mental property on that occasion. An idea like this is evident in the functionalist conception of a mental property as *extrinsically* characterized in terms of its ”causal role”, where what fills this role is a physical (or, at any rate, nonmental) property (the latter property will then be said to ”realize” the mental property in question). (Kim 1992b, p.6)

Now, Kim’s notion that extrinsically characterized causal roles are filled with physical properties might be interpreted to imply some sort of psychoneural identity. After all, there is a notable resemblance between Kim’s view and that of Armstrong and Lewis who state that mediating mental states in input-output relations are actually brain states. However, the case is just the opposite. Armstrong and Lewis considere mental properties as ”first-order properties” which correlate with specific brain states. We have already established that this line of thought inevitably leads to a dead end. Hence Kim is expressing an alternative view of mental properties, which is nowadays widely accepted among functionalists: mental properties are not ”first-order properties” but ”second-order properties”, which consist of having a property with a certain functional specification (see, e.g. Block 1990, p.155).

The idea of mental predicates as second-order properties was first introduced by Hilary Putnam in his article, ”Psychological predicates” (1967). Putnam’s description was formulated in the paradigm of Machine Functionalism but its main claims can be translated into a causal-role account. Let us take pain as an example. The concept of pain can be defined in functionalist terms of having a mediating causal role to bring about winces, groans, and sadness or even depression whenever tissue damage occurs. In other words, ”pain” is a property with a certain functional specification: *a pain state is defined in terms of its typical causes and effects and its relation to other mental properties*. The identity theorists believed that what met this specification was a brain state. However, MR undermined this thesis by proving that there are several, maybe even an infinite number of states that can meet this specification. This is why Putnam suggested that pain was something more abstract than just a brain state. By stating that a pain state is *a functional state*, Putnam argued that ”pain” is not, for example, an excitation of the C-fibres but *a higher-level property of having at least one of those properties that satisfy the functional specification of pain* (1967, pp. 226–227). So, to be in pain is not ”to have a property with the functional specification of pain” but ”to have the property of having a property with the functional specification of pain”.(see Kim 1992b, p.15.)

Putnam’s suggestion has a lot of sense in it. It is a very plausible idea that, for instance, part of the human brain which produces pain sensations (i.e. those neural processes which participate in the physical realization of pains) can be separated as a pain-capable system S_n. Furthermore, it is an uncontested fact that this system S_n has several alternative neural states $N_1,..., N_n$, which all satisfy the functional specification of pain and, hence, can individually carry out the realization of pain sensations. Now, when a person is in pain, the actual situation is such that some first-order physical property (= a neural state or process N_x) of the functional

organization (= system S_p) has occurred and realized the second-order property P (= to be in pain, i.e. to have a property $N_1,..., N_n$ which satisfies the functional specification of pain).

The applications of the concept of functional states and the concept of second-order properties are undeniably major steps of development in comparison to the traditional correlation thesis. However, this view of mental states as functional states does not manage to circumvent MR-related problems altogether. First of all, construing mental properties as second-order properties does not make the fact that mental states have multiple realization bases go away. In case of the pain example, this means that the second-order expression ”the property of having a property with the functional specification of pain” still includes in itself a disjunctive element, the property $N_1 \vee N_2 \vee ... \vee N_n$. It seems that the discussion is thrown back to square one: mental properties are disjunctions of their physical realization bases. And this view, in turn, is now just as unacceptable as it was at the beginning of this consideration.

The second problem is an ontological one. Even if we take for granted, that mental states are functional states (i.e. second-order properties), we really do not know what sort of a relationship they have with their realization bases. When the materialist underpinning, that there is nothing over and beyond instantiations of physical properties and relations, is added to the mix, the case only gets more complicated. What is actually meant by the ”physical realization” of a mental state or the ”instantiation” of a mental property? What is the ontological status of second-order properties? These issues will be handled later in this chapter.

3. Anomalous Monism

Antireductionism has been the dominant and practically uncontested position on the mind-body problem since the late 1960s, at least until now. Hilary Putnam’s Multiple Realization Thesis and the functionalist program it initiated bear a major responsibility for this state of affairs. Putnam’s MR showed a whole new set of problems with the identity theory. Many have thought of them as overwhelming and conclusive evidence of the failure of the reductionist program in general. At the same time those versions of functionalism, which have no reductionist attachments, have enjoyed enormous success and popularity among philosophers of the mind. But there is yet another popular theory of mind, which has made a remarkable contribution to the rise of antireductionism. Although there have been efforts to formulate a functionalist interpretation of this theory (Loar 1981), there are good reasons to believe that this version of token physicalism bears very little resemblence to functionalist argumentation (see McDowell 1985). And this is not simply due to the fact that it does not approve mind-brain identities. What is referred to here is naturally Donald Davidson’s ”Anomalous Monism” (=AM), which he presented in the early 1970s.

In the field of the philosophy of the mind Donald Davidson is an individual and genuine thinker. His inimitable and controversial token physicalism has received much more attention than acceptance. Everyone who has taken serious interest in the modern philosophy of the mind has at some point come across Davidson’s views. Many scholars find some aspects of his work appealing, but most of them (including me) have difficulties

in accepting Davidson's theory in its entirety. Davidson's stubbornness on some issues and his courageously alternative approach have probably contributed to the mixed response.

AM is not taken under consideration here simply for the sake of its value as an intriguing piece of philosophical thought. Davidson's view is closely linked with the central theme of this thesis. One of the purposes of this chapter is to outline Jaegwon Kim's notion of physical realization. AM and particularly Davidson's conceptions of causal relations between the mental and the physical happen to be the starting point for Kim's reflections. Hence I am trying to offer here a brief picture of Davidson's philosophy of mind. I will also point out and dwell shortly on some of its much discussed difficulties.

3.1 Davidson's Proof of the Irreducibility of the Mental

Donald Davidson proposed in "Mental events" (1970) a theory about the relation between the mental and the physical that has become known as Anomalous Monism (=AM). In a nutshell, AM is a version of token physicalism, which holds that mental entities (particular time- and space-bound objects and events that can be described in various nonequivalent and nonsynonymous ways [see Davidson 1969]) are physical entities. The controversial part of AM is that, even though it endorses ontological reduction, it resists conceptual reduction. The consequence of this is that, although mental events are physical events, mental concepts cannot be reduced by definition or natural laws to physical concepts. (see Davidson 1993, p.3)

Davidson's conclusion is supported by an argument that is based on three premises: (1) that mental events are causally related to physical events (=The Principle of Causal Interaction), (2) that singular causal relations are backed by strict laws (=The Principle of the Nomological Character of Causality), (3) and that there are no strict psycho-physical laws (=The Anomalism of the mental). Davidson's argument requires one further assumption, (4) that the mental supervenes the physical (=The Supervenience Thesis). This last premise makes possible the claim that mental events are irreducible to physical events, even though they are in fact physical events.

At first glance, premises (1)–(3) seem to be at best mutually inconsistent: premise (3) denies what premise (1) asserts. Furthermore, especially premises (2) and (3) are difficult to accept without additional assumptions and explanations. In order to avoid misunderstandings and misrepresentations, it is best to get to the core of Davidson's argument and start from the beginning, in other words from the background behind premises (1)–(4) and Davidson's reasons for asserting them.

The Principle of Causal Interaction. This first premise asserts that at least some mental events causally interact with physical events (Davidson 1970a, p.208). There seems to be no confusion surrounding this modest and commonsensical claim and hence no fire either. But if we take a closer look at this premise, it turns out to be a real fire-cracker. What I am interested in here is Davidson's claim that causality may run both ways, from the physical to the mental and from the mental to the physical.

The first possibility, causation from the physical to the mental, is as widely accepted among antireductionists as among epiphenomenalist and reductionist philosophers. Davidson's own example of perception makes a convincing case for the claim: "if a man perceives that a ship is approaching, then a ship approaching must must have caused him to believe that a ship is approaching" (ibid.). Assuming that this holds, we are still left with the possibility of mental causation. Davidson believes that various mental events such as intentional actions, decisions, judgements and changes of beliefs can play causal roles in bringing about physical events. As he writes, "I would urge that the fact that someone sank the *Bismarck* entails that he moved his body in a way that was caused by mental events of certain sorts, and that this bodily movement in turn caused the *Bismarck* to sink" (ibid.).

Davidson does admit that there could be mental events that do not have physical events as causes or effects or *vice versa*. However, even one case of mental causation may be enough to cause AM a great deal of trouble. The reason is that Davidson presupposes that the relation between the physical and the mental is a supervenience-relation, which means that mental events are constituted and determined by physical events. If this is so, it is difficult to imagine how a mental event could have an effect on the very same physical event that constitutes and determines it. Nevertheless, the postulation of mental causation would definitely require that. I will examine this issue more fuller later in conjunction with Jaegwon Kim's notion of physical realization. We will discover that Kim especially has a great deal to say about this subject.

The Principle of the Nomological Character of Causality. This second premise repeats a familiar and generally accepted belief concerning causation; that is, causal relations must fall under laws. However, Davidson gives an extra twist to this conception by adding the claim that events related as cause and effect are subsumed by strict deterministic laws (1970a, p.208). Davidson's notion of "a strict law" is the most important supporting element of AM, and its content provides the basis for premise (3), the anomalism of the mental. Unfortunately, explicating the meaning of "a strict law" is anything but a straightforward task and it requires the clarification of some additional assumptions. In spite of the difficulty, I will try to provide a short but accurate interpretation of the concept.

First of all, Davidson (1970a, p.217) sees "laws" as true, lawlike sentences, which "are general statements that support counterfactual and subjunctive claims, and are supported by their instances". (According to Davidson, lawlikeness is a matter of degree, for the confirmability of laws by positive instances brings in a certain amount of vagueness). Because the laws are linguistic in nature, "events can instantiate laws...only as those events are described in one or another way"(ibid.,p.215). What Davidson has in mind is that events instantiate laws only by satisfying a certain description that itself instantiates the law in question (by instantiating a component of the law). In other words, events are subsumed by laws only under descriptions.

As was mentioned, The Principle of the Nomological Character of Causality states that causally interrelated events are subsumed not only by laws but by "strict" laws. Davidson begins to clarify the conception of a strict law by first making a distinction between two kinds of laws, homonomic and heteronomic:

> On the one hand, there are generalizations whose positive instances give us reason to believe the generalization itself could be improved upon by adding further provisos and conditions stated in

> the same general vocabulary as the original generalization. Such a generalization points to the form and vocabulary of the finished law: we may say that it is a *homonomic* generalization. On the other hand there are generalizations which when instantiated may give us reason to believe there is a precise law at work, but one that can be stated only by shifting to a different vocabulary. We may call such generalizations *heteronomic*. (1970a, p.219)

Davidson's point is simple. A law is either heteronomic or homonomic. If a law is heteronomic, it is unstrict, and if homonomic it is strict. The more difficult question is what kind of law can be regarded as a strict or an unstrict law. In order to crack this puzzle, we must introduce yet another assumption; that is, Davidson's notion of "a comprehensive closed theory".

Davidson's notion of a comprehensive closed theory has no exact formulation. In fact, the looseness, that Davidson has allowed to remain at this point, has had an effect on AM as a whole. This lack of clarity is the major reason why the concept of a strict law has been considered to be obscure, and why it has received negative responses and criticism. For Davidson, comprehensive closed theories are ultimate or ideal theories, which gives us "reason to believe they may be sharpened indefinitely by drawing upon further physical concepts: there is a theoretical asymptote of perfect coherence with all the evidence, perfect predictability (under the terms of the system), total explanation (again under the system)"(1970a, p.219). It should be emphasized that what makes a theory comprehensive and closed is that its vocabulary is closed. A theory can reach perfect coherence and perfect predictability by applying only those concepts included in the vocabulary in which the theory is couched. Davidson thinks that a comprehensive closed theory is an example of what strict laws (i.e. homonomic generalizations) are and should be. Correspondingly, unstrict laws (i.e. heteronomic generalizations) are what they are because they fail to meet the condition of a closed vocabulary. In order to reach an adequate level of explanation and prediction, unstrict laws are always required to extend their vocabulary by adapting additional concepts, for instance, from comprehensive closed theories.

Which scientific theories then are ideal or ultimate, and governed by strict laws? Well, any closed theory will do. The term "closed" is used about a theory, when the events within its domain interact only with other events within the theory's domain. Davidson follows this definition to its logical conclusions and states that, for instance, psychology, biology, or even chemistry are not closed sciences. According to Davidson's view psychology cannot reach the status of a homonomic generalization, for the causal interactions, which take place between mental and physical events, extend the domain of the psychological theory to include also the physical domain. Chemistry and biology are not closed either, because chemical events interact with biological events, and because non-chemical events such as the transmissions of light rays have an effect on chemical events.(1970a, p.224; 1974a, pp.230–231.)

Davidson himself believes that, in practice, closed theories can be found in the physical sciences. According to him, what comes closest to a homonomic generalization is physical theory, which

> ...promises to provide a comprehensive closed system guaranteed to yield a standardized, unique description of every physical event couched in a vocabulary amenable to law. (1970a, pp. 223–224).

Davidson has specified this notion much later in "Thinking Causes" by writing that

> ...what I was calling a law in this context was something that one could at best hope to find in a developed physics: a generalization that was not only ”law-like” and true, but was as deterministic as nature can found to be, was free from caveats and *ceteris paribus* clauses; that could, therefore, be viewed as treating the universe as a closed system. (1993. p.8)

Davidson has taken an even stronger position elsewhere and stated that completed physics is a closed system (1980a, p.241). However, the slightly reserved attitude he has adopted in ”Mental events” is understandable. I seriously doubt that even Davidson's favourite example, physics, could be regarded as closed at its present stage. Considering all the explanatory problems linked with quantum phenomena, it is clear that the levels of ”total explanation” and ”perfect predictability” are not within the reach of physics. At least not for the moment. So, Davidson does have every reason to be cautious in his choice of words.

This is, indeed, the most bizarre part of Davidson's notion of a strict law. There simply are no current scientific theories that could be counted as homonomic generalizations in respect to Davidson's criterion. As we will soon see, Davidson continues to build the whole program of AM on the assumption that there cannot be strict laws linking the mental to the physical or the physical to the mental. The ambiguity in the concept of a strict law gives justifiable grounds to question the validity of the latter assumption: Why should the lack of strict psychophysical laws support AM, if there is no certainty that strict laws exist in the first place?

The Anomalism of the Mental. This third principle makes the claim that there cannot be any strict laws which could govern mental events and hence function as a basis for the prediction and explanation of mental events (Davidson 1970a, p.208). Davidson's conclusion is the outcome of two assumptions: (i) there cannot be strict psychological laws, and (ii) there cannot be strict psychophysical laws. Naturally, both of these assumptions are dependent on the conception of a strict law. However, both rely on this conception in slightly different ways, so it is better to consider them independently.

The reason why Davidson believes there cannot be strict psychological laws is that psychology is not a closed theory (1970a, pp. 223–224). The issue of whether psychology is closed or not is a more complex matter. As we previously learned, The Principle of the Nomological Character of Causality states that laws are description-related: an event instantiates a law only when it satisfies a description that does. So, a strict psychological law is instantiated by the mental description, which instantiates the law in question by instantiating one of its components. The problem here is related to the phrase ”mental description”. I mentioned earlier in conjunction with the topic-neutrality problem, that Davidson thinks of mental descriptions as open sentences that contain a mental predicate. The mental predicate is, in turn, defined as a predicate that contains a mental verb, in other words a verb that expresses a propositional attitude (see 1970a, pp. 210–211). All this adds up to the fact that, in order for a psychological law to be strict, it has to contain a mental predicate; that is, it has to be instantiated by an event which has a mental description.

Now, there is a reason to believe that this cannot happen. Davidson made clear that a comprehensive closed theory is as explicit as possible. Hence the causal laws that govern it must be as explicit as possible too. However, Davidson maintains that mental predicates do not go along with explicit causal laws, because they are *dispositional* (see 1976, pp. 273–275). Brian P. McLaughlin explains Davidson's position by noting, that

functional or dispositional predicates ”do not explicitly state the causal bases for the dispositions they express, the causal bases that bring about certain characteristic effects”(McLaughlin 1985, p.345). Hence, laws containing mental predicates always have to be backed up by some more explicit laws, which lack these mental predicates. Because mental predicates require further support from laws that do not contain functional or dispositional predicates, they evidently cannot be allowed to appear in the basic vocabulary of a comprehensive closed theory. That is why Davidson claims that psychology is not closed, and that there cannot be strict psychological laws.

Event though there cannot be strict psychological laws, there is still the possibility of a strict psychophysical law. The fact that mental predicates cannot be a part of the vocabulary of a closed comprehensive theory does not rule out the possibility that there can be laws connecting the physical and the mental. There is the theoretical possibility, that mental predicates can be replaced with other (i.e. physical) predicates of the basic vocabulary of a closed comprehensive theory with the help of proper bridge laws. If mental predicates are actually reducible *via* bridge laws to physical predicates, they become a part of the general physical vocabulary – or to be precise, a part of the extended vocabulary of a comprehensive closed theory. However, this is only a theoretical scheme. The general opinion among philosophers of the mind is quite unanimously against the existence of linking bridge laws between the mental and the physical. Even if some have defended the existence of such laws, they have not been able to give concrete examples of them.

Davidson denies the existence of bridge laws for his own reasons. Davidson believes there to be a categorical difference between the mental and the physical, which has the consequence that ”events as described in the vocabulary of thought and action...resist incorporation into a closed deterministic system” (1974a, p.230). In ”Mental events” Davidson explains the same absence of strict psychophysical laws by referring to ”the disparate commitments of the mental and the physical schemes” (1970a, p.222).

By ”a categorical difference” or ”the disparate commitments” of the mental and the physical Davidson means that both of these domains have unique constitutive principles. These principles are regulative elements, which govern the application of concepts. For instance, in the case of the physical scheme, constitutive principles regulate the measurement of length, weight, temperature etc. by providing an asymmetric and transitive two-place relation, a framework, within which the measurements are carried out. As an outcome these principles form two-place relations such as longer than, heavier than, and so on (Davidson 1970a, p.220; 1973a, p.254). In the case of the mental scheme, the dominant constitutive principles are, in turn, the principles of rationality. The basic idea is, that when a person’s actions and behavior are evaluated, one usually attributes to him certain propositional attitudes (beliefs, desires, thoughts, opinions etc.) on the basis of which the behavior is interpreted. In order for this interpretation to be correct, these beliefs and attitudes have to be coherent and consistent with each other. In other words, the principles of rationality regulate the application of mental concepts by requiring at least some minimal degree of rational coherence from the content of the person’s propositional attitudes which are engaged in bringing about certain behavior. (see Davidson 1974a, p.236–237)

The reason why Davidson does not approve the existence of bridge laws is that mental predicates need the norms of rationality. Mental concepts always have some rational conditions of application, while physical concepts have none. As was mentioned earlier, in a reduction the reduced concept becomes expressible by the concept into which it is reduced. In this case, it would mean that a mental predicate becomes a part of the

vocabulary of a physical theory. Mental predicates would also need to have nonrational conditions of application, for the physical domain is not governed by any principles of rationality. However, because Davidson believes that mental predicates cannot have nonrational conditions of application, they cannot be reduced via bridge laws to physical predicates. Finally, because there cannot be psychophysical bridge laws, Davidson holds that neither can there be strict psychophysical laws.

The Supervenience Thesis. By denying the existence of strict psychophysical laws and bridge laws Davidson puts himself in a rather difficult position. The anomalism of the mental has two problematic consequences that do not sit comfortably with the rest of the argumentation. First of all, Davidson holds that at least some mental events are causally interactive with physical events (The Principle of Causal Interaction), and that all causal relations must be governed by strict laws (The Principle of the Nomological Character of Causality). Now, Davidson's notion that there cannot be strict psychophysical laws (The Anomalism of the Mental) evidently contradicts the first premise, because there are no strict laws subsuming interactions between mental and physical events. Secondly, the denial of linking bridge laws rules out psychoneural identification and guarantees autonomy for the mental domain. However, it simultaneously leaves the ontological status of mental events open. If further characterizations regarding the antireductionism of mental events are not offered, one has to deduce that Davidson sees the mental as an independent substance and hence proposes some sort of dualistic position. And I am sure we all agree, that the doctrine of dualism is better suited to medieval Christian theology than the modern philosophy of the mind.

In order to fill these gaps in his argumentation, Davidson introduces the fourth and final premise, The Supervenience Thesis. We will later have more time to dwell fully on supervenience and some its problems. Therefore, I will present here only what is necessary to complete our description of Davidson's AM.

The Supervenience Thesis states that mental events supervene physical events. By "supervenience" Davidson means a dependency-relation which holds between two families of properties, even though there are no laws governing this relation. The concept of supervenience originates from moral philosophy and is, therefore, usually illustrated by describing the dependency-relation between moral and non-moral properties. For instance, the moral property of "being a good person" cannot solely be defined on the basis of a person's natural properties (such as height, weight, personal history, environment etc.), but evidently the property strongly relies on these properties. The Supervenience Thesis states that if two persons are identical in respect to their natural properties, they also have to be identical in respect to their moral properties. Davidson holds that mental events are dependent on physical events in exactly the same way. In "Material mind" he explicates this by saying that "it is impossible for two events (objects, states) to agree in all their physical characteristics and to differ in some psychological characteristic" (1973a, p.253). In other words, if two events are identical in respect to their intrinsic physical properties, they necessarily have to be identical in respect to their mental properties as well. Furthermore, because there are no strict laws or bridge laws linking the mental and the physical, mental events cannot be reduced to physical events, though they are dependent on them. (see also Davidson 1970a, p.214)

But what has any of this to do with the problems linked to the denial of psychophysical laws as well as bridge laws that I introduced earlier? The connecting link is that The Supervenience Thesis paves the way for

Davidson's thesis of token identity. As we surely remember, token physicalism holds that every mental event is a physical event, even though it does not endorse reduction between them. As Davidson puts it:

> Anomalous Monism resembles materialism in its claim that all events are physical, but rejects the thesis, usually considered essential to materialism, that mental phenomena can be given purely physical explanations. Anomalous Monism shows an ontological bias only in that it allows the possibility that not all events are mental, while insisting that all events are physical. Such a bland monism, unbuttressed by correlating laws or conceptual economies, does not seem to merit the term "reductionism"...(1970a, p.214)

It is appropriate to recall that token physicalism does not claim that some physical event *P* is identical with some mental event *M*. Instead, it only says that there are events of two kinds, the physical kind and the mental kind.

Now, whatever you might think of token physicalism, Davidson's argument for token identities is sound in terms of the rest of his reasoning. Premises (1) (=some mental events causally interact with physical events) and (2) (=causal relations are always subsumed by strict laws) are not in contradiction with premise (3) (=there cannot be strict psychophysical laws or bridge laws), because they, taken together, imply token physicalism. If mental events causally interact with physical events, they have to fall under strict laws. The only strict laws are physical laws. Hence mental events have to fall under physical laws, and the only way they can do it is by being physical events!

Furthermore, premise (4) (=the mental supervenes the physical) does not only take away the ontological obscurity concerning mental events, but it also supports the conclusion. Mental events are not independent substance-like entities, for they are irreducible, supervenient properties of physical events, which are constituted and determined by physical events. Because constitutive physical events and their supervenient mental events are not independent entities but one and the same thing, it is reasonable to claim that mental events are physical events. Since physical events fall under strict laws, and since mental events are inseparable parts of physical events, it is also natural to assume that mental events could also fall under those very same laws.

Now, we finally have seen the whole picture. Davidson offers a fairly difficult (but convincing) proof for the irreducibility of the mental. The token identity thesis does agree that on an ontological level mental events can be reduced to physical events, because they are one and the same thing. However, the absence of both strict pyschophysical laws and bridge laws avoids the possibility of a conceptual reduction between the mental and the physical. Davidson's argument for AM is sound. This means, that in order to refute his conclusion, one has to try to tackle at least one the premises. Below I will consider some possible approaches that might just do this job.

3.2 Some Debatable Issues

There are numerous excellent papers available on either AM or on some specific aspect of Davidson's theory such as his conception of events, causation, and laws (see, e.g. LePore & McLaughlin [eds.] 1985 or Vermazen & Hintikka [eds.] 1985). I have already briefly pointed out some of the possibly problematic claims related to these issues. However, there is still some space left to add a few additional notes. I am not going to offer any detailed

analysis of the weaknesses in Davidson's argumentation. Neither will I provide meticulous reflections on each of the underlying assumptions in a search for defects. This would not serve the purposes of this thesis. My intention is merely to point out that Davidson's premises and conclusion do have debatable features. In addition, I will indicate a couple of points of departure, from which more thorough counterarguments can be constructed. The issues related to supervenience will be dealt with later in connection with Jaegwon Kim's philosophy.

The Concept of an Event. Davidson's program of AM relies heavily on his conception of an event. In fact, in "Mental events" (1970) Davidson holds that the mind-body problem is not about the seemingly incompatible relations between conscious acts and brain processes or mental states and physical states. Instead, the settlement of this issue requires only the positing of an adequate relation between physical and mental *events*. Furthermore, causal relations – and especially their property of being subsumed by strict laws – play a critical role in establishing the irreducibility of the mental. As already mentioned, according to Davidson, causation itself is a relation between events. So, evidently the plausibility of Davidson's conception of an event is not a minor detail, but a crucial factor in evaluating the plausibility of AM.

Davidson has argued in numerous articles (see 1969;1970b;1971a) that events are concrete particulars, in other words entities which have spatio-temporal locations. Naturally, this is not the only position available on events. Some philosophers – such as N. Wilson (1974), T. Horgan (1978), R. Trenholme (1978), and R.M. Chisholm (1976;1985) – have taken the strict ontological view, which states that there are no events (especially no particular events) in addition to states of affairs. For instance, Chisholm has argued that an event is an abstract, eternal object – a state of affairs – which is not "dependent for its existence upon anything that is not an eternal object"(1985, p.110). Because events (i.e. states of affairs) are not dependent for their existence upon anything that might not have existed, Chisholm concludes that events cannot be particulars. (for a critique of no-event metaphysics, see Thalberg 1985)

Davidson's conception of event can also be criticized from another angle. These counterarguments are not targeted against the nature of events but rather against Davidson's reasons for asserting their existence in the first place. Davidson has given at least three kind of arguments for positing events. In "Causal relations" (1967b, pp. 149–162) he holds that the most plausible interpretation of singular causal statements is to consider them as two-placed predicative statements, where the singular terms designate events. In plain terms, Davidson thinks that causation cannot be explained meaningfully, if events are not posited as the causal relata. In "Individuation of events" (1969) he repeats the argument almost in the same form. Davidson states that the explanation of an avalance does not make sense, if one does not postulate two concrete events (= the cause and the avalance itself), which will instantiate some causal law needed to fulfil the description of the occuring causal relation (ibid., p.165). Now, both of these arguments are weak. In order to refute them, one has only to offer an alternative account of the causal relation, which has no need for the predicate form and event-designating singular terms (see, e.g. Mackie 1974).

The most forceful and best articulated defence for the existence of events are Davidson's semantic arguments. These arguments are not directly linked with AM, for they rather concern Davidson's semantic theory for language (see 1967a;1970d;1973c). Davidson's central argument is, that in order to get the correct truth

conditions and the correct logical forms for action, event, and causal sentences, one has to quantify over entities that stand in causal relations and have spatial and temporal properties. For Davidson these sorts of entities are particular events. (for critical evaluations of the view, see e.g. Horgan 1978; Bennett 1985; Quine 1985)

Causation and Strict Laws. The second premise of AM, The Nomological Character of Causality, states that causally interactive events are subsumed by strict laws. Now, this principle contains some assumptions that can be questioned.

First of all, Davidson holds that causation is a genuine relationship between particular events, not of events under descriptions. This belief is the reason Davidson wants singular causal statements to be interpreted as two-placed predicative statements, which contain singular terms designating particular events (see 1967b, pp.149–162; 1980b, pp.154–155). However, this notion has been contested by P.F. Strawson (1985). He claims that ”there is no single natural relation which is detectable as such in the particular case, which holds between distinct events or conditions and which is identifiable as causal relation” (ibid., p.120). Strawson has clearly taken the side of Kant against Hume by denying the possibility of natural causation and yet endorsing the idea of causal efficacy, which ”is not derived from experience of a world of objects, but is a presupposition of it; or, perhaps better, is already with us when anything which could be called ’experience’ begins” (Strawson 1985, p.128).

Strawson’s example only shows that Davidson’s conception of causation is not immune to criticism. One can either simply accept Davidson’s proposal or adapt the Humean supervenience view, which states that causal relations supervene particulars. However, it is just as justifiable to take any intermediary position between these two extremes. In any case, the chosen ontological view of causation does affect the plausibility of AM – either positively or negatively. (for a comprehensive account of the current debate on causation, see Sosa & Tooley [eds.][1993])

Davidson’s notion of strict laws and laws in general is also debatable. According to his view, particular events instantiate laws by satisfying a description that does. According to this notion Davidson assumes that laws are sentences. This conception of law has been questioned by Fred Dretske (1977). However, I am not going to take up this issue here, for there is a more interesting point of view regarding Davidson’s notion of laws; that is, the comment offered by Jennifer Hornsby (1985). Hornsby has made the very appealing observation that neurophysiological occurrences in the brain are too microscopic to be those particular physical events that Davidson had in mind while formulating AM. Hornsby states that ”someone who thinks that all causally related events can be subsumed by laws formulated by scientists will sometimes claim that events we recognize are in fact fusions of the events described in laws” (ibid., p.457). Hornsby’s suggestion is that if mental events are brought under laws, they have to be related with mereological sums of microscopic processes. These mereological sums are constituted by neural events, which participate in bringing about certain mental states, and which together add up to physical instantiations of these very same mental states. In other words, mereological sums are constitutions of particular physical events. In chapters II and III we will discover that empirical evidence supports Hornsby’s conclusion. But it is an entirely different issue as to whether or not physical events understood as ”fusions of neural events” are any more compatible with token physicalism. After all, a token physicalist

interpretation would state that a mental event is the same thing as several neurophysiological processes, which do not overlap spatially or even temporally. This sort of situation, in turn, is probably the worst position in which to start formulating any kind of identities – irrespective of the metaphysical or ontological convictions one might have.

I have already commented on Davidson's notion of a strict law, which undoubtedly is the most controversial of his claims. Davidson thinks that homonomic generalizations can be found only from comprehensive closed theories. However, what makes this claim so dubious is that current science does not contain or reveal any such theories or nomological necessitations. In fact, many have not been convinced by the postulated need for a closed theory. For instance, Alexander Rosenberg (1985, p.404) does not believe that any science has to start with a closed theory and strict laws. Rosenberg explains his belief by a historical example: there was physics as a scientific practice before Newton, just as there was chemistry before Mendelev. One might add in support of Rosenberg, that there is physics and chemistry even after Newton and Mendelev, even though we do not have closed theories and unimpeachable strict laws. The general point is that science manages and is successful without the exaggeratedly tight criteria of scientific practice. Also Daggfinn Føllesdal (1985, pp.320–321) has presented similar thoughts, by stressing that even open systems, whose components might be influenced by factors outside the system, can be subsumed by strict laws. Furthermore, Føllesdal holds that the mental and the physical, in fact, fall under the scope of one and the same theory, which in addition is a closed comprehensive theory.

The Non-existence of Psychophysical Laws. Davidson's token identity thesis, that mental events are physical events, rests on the premise that there cannot be strict psychophysical laws or bridge laws. The fact that there are some causal interactions between the mental and the physical, in spite of the lack of linking laws, is sufficient proof of token physicalism for Davidson.

The notion of the anomalism of the mental has been particularly difficult to swallow for many philosophers of the mind. This reluctance is largely due to the consequence of the notion – token physicalism – which is something most of us feel too uncomfortable with. Hence, there have been numerous attempts to rediscover the missing psychophysical link. For instance, Ernest Sosa (1993, p.48) has claimed that even though there are no strict psychophysical laws, the mental can be connected with the physical through some relation of modality, conditionality or unstrict laws. Jaegwon Kim (1985, p.199), in turn, has held that there are psychophysical, nonpredictive normative laws and principles, and that this is not in any way inconsistent with the anomalism of the mental. Mark Johnston (1985, p.425) has noted that the denial of psychophysical laws is acceptable only in the light of the interpretative view; that is, that the mental and the physical are thought to be categorically divergent – that they have "disparate commitments" (to use Davidson's term). Then Johnston goes on to argue that the way in which Davidson has postulated this distinction is mistaken, and that there can very well be psychophysical generalizations, though these might be less forceful than strict laws.

These counterarguments form only the tip of the iceberg, and they are presented only as examples. I believe that the most effective argument against the anomalism of the mental and AM in general has been left out.

What I have in mind are the criticisms, which started to appear during the 1980s charging Davidson with epiphenomenalist tendencies. I will return to this issue later in connection with Jaegwon Kim.

The Token-Identity Thesis. Finally, I wish to introduce an argument against the token identity thesis that is formulated by Terence Horgan and Michael Tye (1985). The reason why I want to bring this up here is that the writers have acknowledged a very important aspect of Anomalous Monism that has either mistakenly been regarded as a minor point or silenced to death. Davidson's AM has generally been placed within the antireductionist movement that started to take shape in the late 1960s after Hilary Putnam launched the Multiple Realization Thesis and the functionalist program along with it. Horgan and Tye have discovered evidence that they believe shows that this conception is wrong. What they have actually done is that they have taken Putnam's MR and adjusted it to apply also to event identities, not just to state or property identities. The following argument is the outcome of their reflections.

Horgan and Tye have defined "events" in their broadest sense, according to which they include also changes, states and processes. In addition, they give two definitions, one for the mental domain and one for the physical domain (Horgan & Tye 1985, pp. 427–428):

> [The Definition of Mental Events] For any creature C who has mentality, there is a non-empty set M(C) containing all and only the mental events of which C, at one time or another during his lifetime, is the subject: we shall call this C's *mentality set*. We shall call the contents of M(C) to include not only events of the kind that are apparently posited by common-sense psychology ("folk psychology"), but also mental events of any additional kinds that would be posited by an ideal theoretical psychology.
>
> [The Definition of Physical Events] For any creature C with a non-empty mentality set M(C), we shall say that set of events $P(C)_i$ is a *physical causal isomorph* of M(C) (for short, a PCI) iff (1) every member of $P(C)_i$ is a physico-chemical event of which C is the subject, and (2) there is 1–1 relation R between the events in $P(C)_i$ and the events in M(C), such that (a) each event in $P(C)_i$ is simultaneous with its R-correlate in M(C), and (b) the events in $P(C)_i$ collectively conform to all the causal principles of common-sense psychology and theoretical psychology which govern their respective R correlates in M(C).

Now, the argument itself is a simple one. Horgan and Tye state that every creature C with a M(C) is inclined to have several distinct PCIs for his M(C). Because M(C) has several PCIs, some events in M(C) do not have unique correlates among M(C)'s PCIs and are hence not identical with any of the PCI-correlates. The unavoidable conclusion is, that since an event in M(C) is not identical with any of the PCIs, then the token physicalist claim that mental and physical events are in fact one and the same physico-chemical events is untrue.

When the argument is stripped of its techinal jargon, there is not much originality left. But its ingeniousness lies elsewhere. Horgan and Tye have simple asked two questions: If the multiple realizability of mental states prevents state-to-state and property-to-property identifications, why should it not also prevent event-to-event identifications? And if multiple realization prevents identifications between event-types, why should it allow identifications between event-tokens?

The answer to these questions is largely as follows. Multiple realization is an inseparable part of the human brain function irrespective of whether the mental is described in terms of events, states or other

phenomena. The concept of an event does not in itself contain anything that would change this situation. However, the second question is the more interesting one. How does token physicalism square up with multiple realization? As Horgan and Tye have stated, it fails to do so. I believe that when token physicalism is put alongside with MR, their incompatibility becomes apparent. The reason is that if one does not accept disjunctive properties, neither should one accept any kind of identity between a mental event M and a disjunction of physical events P_1 v,...., v P_n.

It seems that when MR is introduced to token physicalism, the postulated token identities simultaneously turn into triangular or multiple dramas. At first there is only two events, but suddenly the number increases. Now, there are three, four, or maybe even five events that all should be identical with each other. And what makes this all even less acceptable is the fact that Davidson sees events as concrete particulars. Normally, token physicalism would hold that there is one particular event that can be seen as a physical or a mental kind. When multiple realization is added, this notion has to be altered. There is no other way to explain how an event of a mental kind can have so many alternative physical kinds. Either the number of particulars has to be increased or then the number of categories under which an event can fall must be increased. Both of these alternatives are impossible, and I really do not know if they even have any meaning in practice.

These problems support and strengthen the belief that I have had for a long time. Token physicalism is nothing but Spinoza's double aspect in a new guise. The only reason why it has received some acceptance is that it is so vague it does not really mean anything. And because it does not mean anything it allows you to have the autonomy of the mental while at the same time, by accepting ontological reduction, it gives one the chance to flirt with the neural sciences. That is the beauty of it. After all, who would not want to have their cake and eat it?

4. Problems with Reduction

The concept of reduction has been passed over several times in this thesis. A range of arguments and reasons have also been presented either in support or denial of the mind-body reduction. But at no point has there been a detailed description of the concept of reduction or a clarification as to what mind-body reduction is really about. However, I am about to rectify this shortcoming here.

The concept of reduction has a strong bearing on the chosen ontological status of the mental. In fact, it is one of the two decisive factors, which determine whether one endorses or eschews mind-body reduction. The other factor is the nature of the relation between the mental and the physical. Both of these issues are closely interrelated. If reduction in general is thought to hold only when it is governed by powerful laws or bridge laws, then the possibility of a mind-body reduction is either accepted or denied on the basis of whether there actually are any such psychophysical laws or principles connecting the mental with the physical. On the other hand, if one proposes a view that postulates the existence of a strong nomological necessity between mental states and physical states, it gives grounds for suspecting that at least partial or local reductions might be possible. In the latter case, "nomological necessity" does not necessarily mean psychophysical laws. Thus, one might propose a

version of reduction that requires only the existence of a some sort of logical bond between the mental and the physical – of course, provided that it is strong enough to fulfil its function.

In the following pages I will try to shed some light on the concept of reduction. I will also consider how it has been applied to the special case of mind-body reduction and how it is related to the debate on the mind-body problem. The major positions on the relation between the mental and the physical are also examined. I am going to end this chapter by addressing some of the main problems linked with the idea of mental reduction. I will also summarize the objections that have previously arisen in our discussions.

4.1 Ernst Nagel's Inter-Theoretic Reduction

The current debate on mind-body reduction and reductionism in general was originally launched by Ernst Nagel in the 1950s. Nagel (see 1961) formulated an account which stated that reduction was a relation between two scientific theories, in which the target theory was reducible to the other theory which, in turn, functioned as a reduction base or a base theory. In Nagel's model theories are understood as collections of laws expressed by statements. These basic laws or axioms constitute the grounds for each theory, which is then supplemented by all statements that are logically or mathematically derivable from them. Each theory also has a characteristic vocabulary formed from non-logical, descriptive expressions such as "temperature", "mass", "length", "gene" and so on. The set of statements which contain the basic laws of the theory are formulated in terms of these expressions. (see Kim 1996, pp. 212–213)

Nagel's inter-theoretic reduction can be illustrated with a simple example. Say we have a biological theory, a target theory T_B, that will undergo reduction to a chemical theory, a base theory T_C. According to Nagel, an inter-theoretic reduction is basically a reversed *logical derivation* or *provability* of one theory from another. This view holds that if a reduction is successful, it shows that the reduced theory is not an independent theory but only a subtheory of the reducer. To put it more explicitly, in order for theory T_B to be reduced to theory T_C, all the laws of T_B must be derivable from the laws of T_C. Thus, all the laws of T_B are theorems of T_C, which should be provable from the set of basic laws of T_C.

There is a clear problem with this approach. Despite having two theories, T_B and T_C, to deal with we also have two distinct vocabularies, V_B and V_C. It is highly unlikely that any of the descriptive expressions contained in these disparate vocabularies are compatible with one another. Hence it is impossible to derive any of the expressions of T_B from the expressions of T_C. That is why we need the aid of additional premises – "bridge principles" – which help us to correlate the expressions in those two vocabularies. Say we have a T_B -law of the form:

(1) $A \_ B$

which simply states that for anything x, if x has the property A, x has the property B. Now, the expressions "A" and "B" are not part of the vocabulary V_C, which makes it impossible for us to derive (1) from chemical theory T_C. However, with the help of appropriate bridge principles we can find the right correlates from the theory T_B and,

furthermore, form a connection between these expressions. For instance, bridge principles of the following biconditional form would suffice to carry out the derivation:

(i) $A \_ A'$
(ii) $B \_ B'$

which state that for anything x, x has property A iff x has the property A', and correspondingly that for anything x , x has B iff x has B'. The expressions "A" and "B" naturally designate predicates of theory T_C. Now, that the link between both vocabularies has been established *via* bridge principles (i) and (ii), it is possible to formulate T_C - statement (2) from which the T_B -law (1) is derived:

(2) $A' \_ B'$

which states that for anything x, if x has the property A', x has property B'.

Nagel's model shows that bridge principles have a key role in reduction. These principles are in fact indispensable; without them it would not be possible to carry out the whole manouver. However, there are a few points that must be addressed regarding them. So far, we have been using the term "bridge principle", which leaves the status of these auxiliary premises open. Are we dealing with actual laws or something less forceful? If we stick to Nagel's original intention where the reduced theory is only a derivation of the reducer, bridge principles could be considered as mere definitions. In this case, bridge principles would guide the substitution of T_B - expressions with T_C -expressions. The second possibility is that bridge principles are actual empirical correlation laws which establish the connection between the properties of theories T_B and T_C. If radically interpreted, the first option could be considered to deal merely with analytic relations between expressions. The second option could, in turn, be interpreted to certify correlations between genuine properties. However, in reality neither of these extreme cases have to be true. For instance, the case of temperature is a well known intermediary example. The correlation between "temperature" and "mean kinetic energy" started as a reduction warranted by empirical bridge laws, but now that the reduction has become so well established the relation has taken more the form of a definitional substitution.(Kim 1996, p.214)

Another issue related to the status of bridge principles is the question of how strong these principles should be. In the previous example the bridge principles (i) and (ii) were presented as biconditionals, but in some cases – as Nagel himself thought – a conditional form will also suffice. Yet again the precise form of the bridge principles is dependent on the strengths of the derived laws and premises. However, our interest in reduction is dictated by the need to discover whether or not reduction can be adjusted to the relation between the mental and the physical. In case of the mind-body relation, there is a general agreement among philosophers of the mind that psychophysical bridge laws have to be biconditional in form (see Beckerman 1992, p.107).

The first reason for this is to ensure reduction. For instance, in our example the biconditional bridge principles allow us to rewrite the reduced theory in the vocabulary of the reducer by certifying the replacements of the expressions "A" by "A'" and "B" by "B'". Now, if the case is such that T_B is not derivable from T_C, T_B could still be added to T_C as an additional law. After all, it is a true T_C - statement, which, in addition, is in the form of a law. In other words, the usage of biconditional laws guarantees that both derived laws and additional new laws

can be formulated using the basic vocabulary of the reducer without a need for postulating any new properties or entities.

The second reason is already familiar to us from the section that considered the identity theory. As was mentioned earlier, identity theorists proposed psychoneural identifications between, say, mental states that fulfilled the causal role for bringing about pains and physical states such as activations of the C-fibres. Now, the claimed identity of the form ”pain = C-fibre excitation” cannot be carried out unless there is some principle connecting or correlating mental and the physical states with one other. Identity theorists, such as Armstrong and Lewis, thought that the correlation was to be found on the basis of empirical scrutiny. However, they were wrong; psychoneural identification requires a great deal more than just a straightforward correlation.

An identification is an ontological reduction where a higher-level mental property (as well as higher-level facts concerning it) is reduced to a physical property. This sort of ontological simplification requires that, for example, the mental state M (that has the causal role for bringing about pain) and physical state P (i.e. the activation of C-fibres) are connected with a biconditional bridge law $M \_ P$. Now, if the Nagelian reduction is carried out with conditional bridge laws of the form $M \_ P$ or $P \_ M$, the possibility of an identification $M = P$ is ruled out. In the case of the mind-body relation, this means that the whole idea of reduction gets watered down. After all, the aim of establishing a mind-body reduction was originally (and still is) initiated by the need to make some sense of the relation between the mental and the physical. And the sought after result is the ontological reduction of mental states to neural events and processes guaranteed by a successful inter-theoretical reduction mediated by appropriate bridge laws.

One of the sources that inspired discussions of a Nagelian mind-body reduction in the 1950s was Wilfrid Sellars’ article ”Empiricism and the philosophy of mind” (1956). In this paper Sellars suggested that so called ”folk psychology”, which based the explanation of behavior on our commonplace psychological experiences (such as beliefs, desires, expectations, thoughts etc.), could be considered to form a semantically coherent system – a language – amenable to theoretical descriptions. Naturally, this alleged theory-nature of folk psychology aroused an interest in examining whether inter-theoretical reduction could be tailored to relating a mental theory (describing mental states and actions) and a physical theory (describing neurophysiological processes in the brain).

One of Sellars’ students, Paul M. Churchland, has presented a model of the Nagelian mind-body reduction that is applicable in a situation where we can actually deal with both sides of the equation, a complete theory of mental action and a comprehensive neural theory. In such a case, Churchland (1979, pp. 81–83) claims that the old psychological theory T_p can be reduced to or replaced by the new neural theory T_n, if

(I) the reduction contains basic instructions as to how T_n replaces T_p,

(II) the set S_o, which contains the basic assumptions (e.g., background principles and beliefs) behind the reduction, does not have to be radically modified,

(III) the basic structure of T_n contains an image of T_p in a way that S_o is T_p’s image in T_o.

The first condition repeats the requirement of bridge laws or corresponding rules. Churchland does not explicitly state that bridge laws have to be biconditional in form; that is why he also uses the term ”corresponding rule”. However, he is determined that these bridge principles or rules have to connect states and properties of both theories in a one-to-one fashion. In practice, this means that for every mental action or state there has to be a correlating brain state or a process. The second condition is meant to ensure that the psychological theory cannot be replaced by the neural theory if the latter is not appropriate. In other words, if the state of neural sciences is not evolved enough or if its theories are utterly wrong, the set of underlying assumptions (S_o) has to be rewritten to fit the requirements of the psychological theory. In this case, one has drifted too far away from Nagel's original intention, where the target theory was supposed to be derived from the base theory. Thus the reduction has to be abandoned. The third condition, in turn, claims that the target theory cannot be replaced by any arbitrarily chosen theory. It also defends the status of reduction as something more than just a re-definition of concepts. The basic idea is that the set of underlying assumptions (S_o) is the only framework within which the correspondence between the properties of the two theories and the linking bridge laws can be found and determined.

In conclusion, I wish to note that in the following pages we will find out that the most difficult problem with Churchland's model (or with any other version of mind-body reduction) is not that we do not have a complete psychological theory or a comprehensive neural theory. There are probably even graver problems that originate from the various positions held on the relation of the mental and the physical; that is, there are views that regard the autonomy of the mental as an ontological fact which cannot be overruled even with existence of bridge laws. In the following I will consider the relation of the mental and the physical in detail and show what makes the mind-body reduction such a special case.

4.2 On the Relation of the Mental and the Physical: From Emergence to Supervenience

According to Ilkka Niiniluoto (1994, p.38), materialism can be divided into three forms on the basis of attitudes towards the mind-body relation. The first form is *eliminative materialism*, which played a minor role in the 1960s' mind-body debate, but which was revived in the 1980s by Stephen Stich (1983) and the Churchlands, Paul M. (1979;1981;1984) and Patricia Smith (1986). Eliminative materialists hold that folk psychology is an utterly false theory that should be rejected and be replaced by a new neuroscientific theory. The second form is *emergent materialism* proposed by K.R. Popper and J.C. Eccles (1977), Joseph Margolis (1978), Roger Sperry (1980), Mario Bunge (1980), and Thomas Nagel (1986). This view takes mental states to be genuinely new, higher-level properties of complex systems, which are causally efficacious. This group also includes all the other forms of non-reductive materialism (or physicalism, if one prefers), which do not endorse the mind-body reduction. The third form is *reductionist materialism*, which claims that mental phenomena can be reduced to physical states and processes. For instance, the identity theorists of the 1950s and 1960s – Feigl, Place, Smart, Armstrong and Lewis – could be seen as proponents of this view.

From the perspective of these three options, eliminative materialism falls outside the scope of our investigation. The primary concern here is to study what sort of a relation can or should be postulated between

the mental domain and the physical domain. Since eliminative materialism denies the existence of consciousness and mental states as described in folk psychology, there is no relation left to be considered. However, all the versions of reductive as well as non-reductive materialism are very much of interest to us. The reason for this is that since the early 1970s almost all of them have referred either to the concept of emergence or to the concept of supervenience in explaining the mind-body relation. During the mid 1980s the focus swiftly shifted to the concept of supervenience, and since then supervenience has retained its position as one of the most discussed topics in the philosophy of the mind. However, during recent years published comments on supervenience have been mostly critical and pessimistic in nature.

In the following pages I will carefully examine both emergence and supervenience. The backgrounds and the meanings of these concepts will be presented; their strongest points and most damaging flaws will be considered. I will also try to shed some light on what their exact role in the mind-body debate might be, and whether these concepts endorse Nagelian reduction.

I

(1.) Interest in the concept of emergence was revived in the 1970s, when it was introduced as a possible solution to the mind-body problem. The contributions of Popper, Bunge, and Sperry especially had a major role in shaping the idea that emergence might be the right way to state dependency-relations between mental and physical states. However, the historical roots of the concept of emergence originate in the work of the nineteenth century philosophers John Stuart Mill (1843) and G.H. Lewes (1875). Both scholars were investigating cases of causation in which several causes participated in bringing about a certain effect. They distinguished between two different event-groups and made the following observations (see Stephan 1992, p.28):

> 1. The effect of the causes acting together is the algebraic or vectorial sum of the effects each cause would have had if it had acted alone.
>
> 2. An effect that is brought about by several distinct causes together cannot be reduced to the sum of independent effects brought about by individual causes.

Mill called the causes of the first group ”homopathic” and the causes of the latter group ”heteropathic”. Lewes also accepted the very same distinction, but he referred to the first type of causes as ”resultants” and the latter type as ”emergents”. Lewes is the first person known to have used the concept of emergence in a technical sense.

The concept of emergence has been criticised as being a somewhat obscure notion. In many cases these criticisms have not been unfounded. There is a certain discernible sense of loosness in Mill’s and Lewes’ formulations. For instance, both philosophers share the belief that heteropathic or emergent effects cannot be *predicted* before their first occurrence. In other words, they hold that emergent effects cannot be deduced from individual members of the complex cause, because adequate knowledge of all the single cause-factors is not available. Lewes makes this point more explicitly; he claims that if all the aspects and steps of a certain process

cannot be followed and explained, the product of the process must be called emergent. Lewes gives a particular example of the passing through of water and hydrogen, where they quit the gaseous and assume the liquid state. Lewes thinks that as long as this process cannot be explicated in a mathematical formula, the water must be regarded as emergent (see Stephan 1992, pp. 28–29).

Both of these cases can be interpreted to imply *an epistemological notion of emergence*. This notion states that emergent properties or effects are not genuine properties or effects of a complex system or a complex cause; they have only to be considered as such because the present state of knowledge and present theories cannot provide a full explanation of them. However, this does not exclude the possibility that sometime in the future – if science has evolved – the required theories might be discovered and the sought after explication could be reached. In this case, the notion of emergence could be rejected, for it would have become redundant.

(2.) Mill and Lewes saw emergence as non-additivity: there were effects brought about by complex causes that could not be explained by simply adding together all the separate effects of the individual causes. Neither of them thought these sorts of effects to be heteropathic or emergent because they were something totally new, different, or of a higher level in respect to elements of a complex cause. Instead, they were inclined to favour the epistomological interpretation of emergence, which stated that these effects were ”emergent” because present knowledge and theories could not provide an explanation of them. However, a new approach was developed in the early 20th century, which took just the opposite position on the notion of emergence. Samuel Alexander (1920) and C. Lloyd Morgan (1923) defended the so called ”theory of emergent evolution”, which saw emergence as novelty; that is, the concept of emergence was used to characterize complex structures that formed genuinely new higher ontological levels. Thus emergence was not considered to be an epistemological description but an ontological concept that designated higher levels of existence.

Early twentieth century emergentism went beyond simply presenting a scientific world view. Its approach was rather cosmological. The world was not depicted merely as an evolutionary process; the theory was supported by a strong assumption that the world was divisible into structured layers; that is, that the world was a hierarchically organized system in which each level is more complex than the one below but also dependent on it. For instance, Alexander (see 1920) distinguished four major levels of existence, in other words four levels of emergence: a) the first was matter itself, which emerged from space and time; b) the second was life, which was the result of complex material processes; c) the next was consciousness, which emerged out of vital processes; and d) the last level was that of deity, which was thought to emerge out of consciousness.

The practical interest of the proponents of emergent evolution was to find out how simple factors and particles can form complex systems the properties of which need not (or even cannot) be explained in terms of the micro level. This task was carried out by formulating the notion that properties are emergent, if they are new in respect to the properties of the particles of the lower, microscopic level. Mario Bunge has offered a modern definition of this idea in his ”Emergence and the mind”:

Let *P* be a property of a complex thing *x* other than the composition of *x*. Then

(i) *P* is *resultant* or *hereditary* if *P* is a property of some components of *x*;

(ii) otherwise, i.e. if no component of *x* possesses *P*, *P* is *emergent*, *collective*, *systemic*, or *gestalt*. (1977, p.502)

Bunge's definition makes clear the requirements that have to be met in order for a complex system to have emergent, higher level properties. The key issue is that if a system has new, higher level properties, these properties cannot be found among the properties of the components that constitute the microscopic level. A property is genuinely new and of a higher ontological level only when its first appearence happens on the higher, emergent level of the complex system. And this is something that is not apt to change, even if the development of sciences increases our knowledge of the subject and provides a new set of more accurate theories.

(3.) Probably one of the most influential and useful versions of emergence is presented by C.D. Broad in his famous *The Mind and its Place in Nature* (1925). The book is based on a series of lectures Broad held at Trinity College during the semester and spring terms in 1923. In his work Broad set out to develop a theory which could explain the difference between organisms and nonvital entities without falling into either vitalistic nor mechanistic interpretations. Broad rejected vitalistic notions on the basis that he believed that the behavior of any organism was ultimately determined by its microstructure. Hence the behavior of any system was explainable without references to some postulated, immaterial and invisible entities (1925, pp. 56–58). Broad did not endorse purely mechanistic theories either. He was convinced that even though the system's behavior was determined by the properties and structure of its components, it was not deducible from the laws concerning the parts of the system –whether they were in isolation or in other relations not found in the actual system (1925, p.59).

As a solution, Broad offers his own conception of emergence. An outline of Broad's theory can be found from the following passage:

> Put in abstract terms the emergent theory asserts that there are certain wholes, composed (say) of constituents A, B and C in relation R to each other; that all wholes composed of constituents of the same kind as A, B and C in relations of the same kind as R have certain characteristic properties; that A, B and C are capable of occuring in other kinds of complex where the relation is not of the same kind as R; and that the characteristic properties of the whole R(A,B,C) cannot, even in theory, be deduced from the most complete knowledge of the properties of A, B and C in isolation or in other wholes which are not of the form R(A,B,C). (1925, p.61)

Broad illustrates the theory by giving a concrete situation in which his notion is applicable. His example is water, a chemical compound formed by hydrogen and oxygen. Broad states that any knowledge we might have of oxygen or of elements other than hydrogen with which oxygen is able to form compounds, does not give us any reason to suspect that oxygen could combine with hydrogen in the first place. The same goes for hydrogen also well. Broad also adds that most of the chemical and physical (qualitative or quantitative) properties of water have no connection with the corresponding properties of oxygen or hydrogen. In plain terms, information about the properties of two distinct components or even knowledge of the effects these components cause when forming combinations with other constituents (than each other) do not facilitate the prediction of the properties of water. This is the reason why Broad thinks water should be considered as emergent. (1925, p.63)

It might be difficult to see in what way Broad's formulation diverges from the other notions of emergence that we have so far seen. Or is it even divergent? In fact, Broad does have a very unique view on emergence; our

considerations have been so general in nature that the essence of his theory has not yet been examined. However, Achim Stephan has presented an excellent explication of Broad's position, which will undoubtly help us to correct this situation. Stephan proposes the following:

> Let S be a system, composed of constituents $C_1,\ldots, C_n$ with the microstructure $[C_1,\ldots, C_n; O]$ (that is, the constituents $C_1,\ldots, C_n$ stand in relation O to each other).
>
> Def: A system property P is called emergent iff
>
> (a) there is a law P_L which holds: for all x when x has the microstructure $[C_1,\ldots, C_n; O]$ then x has property P.
>
> (b) P_L cannot be deduced from laws concerning the $C_1,\ldots, C_n$ in isolation or in other microstructures, even together with compositional principles. (1992, p.37)

Stephan's example reveals the core of Broad's notion: Broad sees emergence as nondeducibility. According to Broad, there is no way we can deduce the law P_L from the laws or compositional principles concerning the microstructure $C_1,\ldots, C_n$.

But how strong is Broad's refutation of deducibility? There are two answers. On some occasions, Broad stresses that some processes are "absolutely emergent", which means that their emergent laws cannot ever be deducible no matter how scientific theories might change and evolve over time (see 1925, p.63). This has the consequence that the emergent properties of these sorts of processes are also absolutely emergent: they are absolutely genuine and undeducible properties. This claim can justifiably be interpreted as supporting the ontological notion of emergence. However, Broad states elsewhere just the opposite; he thinks that it might be logically possible that emergent laws (and hence also properties) appear only "due to our imperfect knowledge of microscopic structure or to our mathematical incompetence" (1925, p.81). Which of the two positions does Broad actually propose?

This is not an easy problem to settle; we could just as well go either way. Firstly, Broad's arguments for emergent properties are not entirely convincing. For instance, I am not sure that chemists would any longer agree that the properties of water cannot be predicted on the basis of the knowledge we have of its components, hydrogen and oxygen. There have been suggestions implying that chemistry could be reducible to physics. If this were actually so, it would mean that the prediction could be carried out: the knowledge of the properties of the constituents provided by physics would suffice to determine the properties and behavior of the compound formed (see Pihlström 1996, p.262). So, evidently there are reasons to favour an epistemological interpretation of Broad's theory. However, I believe that Broad's notion should not be taken out of context. It should be remembered that *The Mind and its Place in Nature* is above all a book about the philosophy of the mind. If we accept the epistemological interpretation, which states that emergent properties might be deduced in the future from the structure and properties of the microstructure, Broad could be classified as a reductionist. After all, to say that an emergent property is *deducible* from a microscopic property (or from some set of microscopic properties and structures) is to say that that emergent property is *reducible* to that microscopic property.

But how would Broad himself react to the accusation of being reductionist? I wish to note two things: Firstly, Broad devotes a lot of time and energy in *The Mind and its Place in Nature* to proving that if

consciousness was considered as an emergent property of the brain, that would not contradict the laws of physics (such as the law of preservation of energy; see 1925, p.109). Secondly, I already pointed out that Broad rejects all kinds of mechanistic theories. In the light of this latter remark, I doubt that he would have been any happier with a physicalist theory. Thus my conclusion is that Broad holds emergence primarily as an ontological concept. To characterize him as a physicalist or a reductionist is simply mistaken. However, to accuse Broad of being obscure and careless is probably accurate. He is not however the first emergentist to fall into this trap.

(4.) The grounds for emergence and also its general features were mostly established by early writers such as J.S. Mill, G.H. Lewes, S. Alexander, L. Morgan and C.D. Broad. These philosophers are thought to form a tradition known as ”British Emergentism”, which has had modern followers at least in the form of Roger Sperry (1980) (for more on British Emergentism, see McLaughlin 1992). Sperry’s views especially have been highly influential among practising neurophysiologists and other neuroscientists, who have very little time or interest in contemplating all the difficult metaphysical issues related to consciousness. However, there are also other more recent views on emergence that are worth mentioning. A number of these follow.

C.G. Hempel and P. Oppenheim (1948) were the first to explicitly state the theory-relativity of emergence. As we saw earlier, the works of Lewes and Broad contained at least some remarks pointing in this direction, although Hempel and Oppenheim were the ones who brought things out into the open. According to them, emergence ”is not an ontological trait inherent in some phenomena; rather it is indicative of the scope of our knowledge at a given time...what is emergent with respect to the theories available today may lose its emergent status tomorrow”(ibid., p.263). So, they both hold that emergence is not an ontological concept but an epistemological one, which is relative to our knowledge of the world. Similar thoughts were presented by Ernst Nagel, who claimed that a property could be emergent only in relation to ”one theory or body of assumptions” but fail to be that in respect of some other theory (1961, p.369). Both of these positions could be summarized by stating that P is an emergent property of x, if we do not know any theory by means of which P(x) could be deduced from our knowledge of the world (see Pihlström 1996, p.264).

Naturally, all the later emergentists, who have considered emergence in the light of Samuel Alexander and other emergent evolutionists, fiercely object to an epistemological interpretation. To them, the concept of emergence will always be an ontological one and emergent properties will remain genuine, novel, and nonreducible properties. Mario Bunge definitely belongs to this category with his *Scientific Materialism* (1981). Bunge states that all true material objects are either systems or components of some system, and that the composition of a system is defined as the set of its parts (1981, pp. 25–27). In other words, to be an emergent property E is to be a property of a system x, which none of its micro-level L_{mic} component parts have. Naturally, this leads to a position, where every system is supposed to possess at least one emergent property: components of x can be systems themselves on the level L_{mic}, but the property E (= being a system on the macro-level L_{mac}) is something that the components of L_{mic} can never possess. (for Bunge’s practical applications of the concept of emergence to the mind-body problem, see Bunge 1980)

Although Bunge does establish the ontological independence of emergent properties (on each level every emergent property is novel and nonreducible in respect to lower levels), his view is rather superficial in nature.

For instance, Ilkka Niiniluoto criticises Bunge's definition of emergence for being too broad: it designates only the properties that belong to a whole but not to its parts. If a system consisted of five parts, we might state that the system has the emergent property of "having five parts". According to Niiniluoto, this is unacceptable, because the definition fails to explicate the idea that the whole is more than a sum of its parts (Niiniluoto 1994, p.43, footnote 8; see also Pihlström 1996, p.263).

Paul Teller has also given a fine analysis of the concept of emergence, in which he considers the same problems that Niiniluoto points out. Teller 's view is similar to the one proposed by Niiniluoto; that is, that an emergent property of a whole should somehow "transcend" the properties of the parts. In order to defend this intuition, Teller offers an excellent guideline, which states that "a property of a whole is an emergent property of the whole when it is not reducible to the NON–RELATIONAL properties of the parts"(1992, p.141). His own example of a property that fails to meet this demand is the emergent property of "being the longest pencil in the box". Naturally, the property of "being the longest pencil" is a relational one. But if we measure all the pencils, we are dealing with a collection of pencils, the properties of which (such as the red pencil being the longest one) are set by the properties of the objects in the collection. In this case, the properties "being red" and "being 10 cm long" are no longer relational, and we are dealing with a situation, where we can reduce the emergent property "being the longest pencil in the box" to non-relational properties of the parts.

We may very well apply this method to one of Bunge's examples. If we have a whole, a heap of five stones, which has the emergent property of "including five stones". The property of "being a stone" is in no way relative to the other stones; we may very well state that a stone is an object of a certain size, a certain shape, a certain material etc. By shifting our attention from the fact that we have a heap of stones, which includes five pieces, to the fact that we have a collection of stones with certain non-relational properties we are on our way to reducing the emergent property "to include five stones" to the properties of the parts (see Teller 1992, p.145). This demonstrates that Bunge's definition does not meet the intuitive requirements set for the concept of emergence; emergent properties defined in this way are not novel, "transcending" and non-reducible properties of the whole.

The last version of emergence that I am going to present here is the view proposed by Karl Popper. Popper's notion of emergence is much too vague to be of interest. Hence I discuss it only briefly. Popper ascribes two meanings to the concept of emergence: 1) emergence is the nondeducibility of properties; 2) emergence is the unpredictability of events. In his book *The Self and its Brain* (1977), written together with John Eccles, Popper gives explanations of both these notions. The first is due to Popper's view of emergent materialism, which claims that during evolutionary development emergent processes "lead, not gradually but by something like a leap, to a property which was not there before" (ibid., p.69). Popper believes that, for instance, consciousness is an outcome of such complex material processes and organizations. The second notion is due to Popper's general belief that the physical universe is essentially indeterministic (see 1966). Popper links indeterminism with emergence by stating:

> ...emergence of hierarchical levels or layers, and of an interaction between them, depends upon a fundamental indeterminism of the physical universe. Each level is open to causal influences coming from lower *and* from higher levels. (1977, p.35)

Now, when these two notions are combined, we are able to formulate Popper's view of the role that emergence plays in the mind-body relation. He believes that properties are emergent in a sense that due to indeterminism atoms can form unpredictable arrangements which lead to physical or chemical properties not derivable from atomic theory. In other words, Popper does not see emergent properties as substance-like entities; they are novel properties of complex material systems. On the other hand, these properties of complex structures are also nondeducible from the components of the micro-level – they cannot be reduced to the structure itself. In plain terms, Popper's position is that consciousness has evolved from complex evolutionary processes, which have given the brain the ability to have emergent properties – mental states – which cannot be explained by or reduced to neurophysiological processes. Nevertheless, mentality is dependent upon occurrences in the brain.

Poppers' conception of emergence itself is vague. Also most of his assumptions and claims starting from the supposed indeterminism of the physical domain can be debated (see e.g. Earman 1986; Niiniluoto 1987). There are nonetheless good reasons for bringing him into our discussions. By openly defending determinism Popper is underlining that even though mental states are determined and dependent upon the physical microstructure, consciousness can still have causal powers in respect of other mental states and the physical domain. I will demonstrate later that this assumption is one the most difficult problems related to the mind-body debate. We are also going to discover that, for instance, Jaegwon Kim does not believe that mental causation is in any way a proof of the irreducibility of the mental but, instead, a reason to endorse it.

(5.) We have so far come across several distinct conceptions of emergence. Emergence has been conceived at least as "novelty and irreducibility", as "nondeducibility", and as "nonpredictability". In addition, we have learned that emergence can be intepreted either as a theory-relative notion, which is dependent upon our knowledge of the world and upon the existing theories, or as a truly ontological concept. This variation does cause problems for our general discussion of emergence, for many of the concepts are mutually exclusive. However, there is a way to circumvent these difficulties. All the versions of emergence share some general features, which can be used to formulate a notion of emergence or a set of principles, common to all the variations.

Jaegwon Kim has proposed that even though there is no unanimous and specific formulation of emergence, it can generally be characterized by the following three theses (Kim 1992a, pp. 122–124; see also 1996, pp. 227–228):

> (1) [Ultimate Physicalist ontology] There are basic, nonemergent entities and properties, and these are material entities and have fundamental physical properties.
>
> (2) [Property Emergence] When aggregates of basic entities attain a certain level of structural complexity ("relatedness"), genuinely novel properties emerge to characterize these structured aggregates. Moreover, these emergent properties emerge *only* when appropriate "basal" conditions are present.
>
> (3) [The Irreducibility of Emergents] Emergent properties are "novel" in that they are not reductively explicable in terms of the conditions out of which they emerge.

To put it briefly, the first thesis notes that the concept of emergence does not include the assumption favored by emergent evolutionists, which claims that emergent levels keep evolving hierarchically from the lower levels *ad*

infinitum. Instead, it is a reality constituted of material, non-emergent entities and their properties (such as mass, size, energy etc.) that functions as the basis for emergent properties. The second thesis, in turn, states that novel properties appear *necessarily* when the complexity of the basal structure meets the conditions required for emergence. ”Necessary” means here at least physical or a kind of nomological necessity but not necessarily an actual law. This second thesis also stresses that emergent properties belong to the whole composed of basic particulars – there is no need to postulate the existence of any new substance-like entity. Finally, the third thesis explicates the intuitive idea related to the concept of emergence; that is, emergent properties are properties of the whole and in this sense more than just the sum of its parts. Hence emergence always implies antireductionism. The justification for this assumption is quite conventional. Firstly, emergent properties cannot be reduced to the basal structure because there are no actual laws linking them with the properties of the basic particles. The other line of thought goes on to prove further that emergent properties cannot be deduced from the basal structure because there are no bridge laws by which a theory of the emergent level could be derived from the law or from laws concerning basic particles and their properties.

On the basis of these characterizations, it is possible to underline the two major tendencies that dictate all versions of emergence. The first is, of course, the irreducibility of emergent properties, the issue of which we have already discussed. The second issue is the more crucial one; that is, emergent properties are *dependent upon* and *determined by* the basal structure. The slogan ”dependence without reduction” neatly encapsulates this central tenet in the concept of emergence. The reason why the dependency-requirement is of so much importance is that emergent properties are supposed to be *explainable* in terms of the basal structure. In fact, the concept of emergence was initially developed for the very reason that phenomena such as consciousness could be explained in the spirit and with the methods of science.

However, emergence has suffered throughout its history from a constant straying between these two interlinked aspects. An adequate explainability of emergent properties can be achieved only through such a dependency-relation in which the properties and structure of the base particles determine emergent properties. Kim’s statement, that emergent properties have to appear *necessarily* when the appropriate basal conditions are present, shows that the dependency in question has to be very strong in nature. But on the other hand the concept of emergence is supposed to guarantee the irreducibility of novel, emergent properties. This, in turn, forces one to avoid postulating over-tight lawlike regularities between base properties and emergent properties, because such notions have the defect of paving the way for reduction.

In summary, the concept of emergence can retain the irreducibility of emergent properties only at the expense of explainability and *vice versa*; an increase in the degree of explanation always simultaneously weakens the ontological status of emergent properties. Both options are equally unappealing. In the first case, we can conceive emergent properties as novel and irreducible but gain nothing in respect of explainability by doing so. For instance, to say that mental states are irreducible emergent properties of the brain in the strong sense does not help us understand the workings of the mind any better, because we cannot benefit from the knowledge provided by the neural sciences. The link between the mental and the physical is too weak to justify any kind of correlation between these two domains. In the second case, we can postulate strong dependency-relations between, say,

certain mental states and some specific neural processes and find ourselves on the way to losing the idea that consciousness is something more than just series of brain states.

Obviously, these examples are highly exaggerated and they offer too negative a picture of emergence. Nevertheless, they do point out the major pitfalls of the concept, which ultimately led to its disappearence from the current mind-body debate. During the last two decades discussions have concentrated on the concept of supervenience, which has been thought to carry the ”dependency without reduction” - idea with greater success.

II

(1.) The concept of supervenience originates in moral philosophy. G.E. Moore has generally been credited with inventing the idea of supervenience in his *Philosophical Studies* (1922), where he describes a certain dependency relation between moral and nonmoral properties. Moore claimes that things that have some intrinsic value or property must possess it invariably under all circumstances, and that any other similar thing must also possess that very same intrinsic value no matter what the circumstances might be (see 1922, p.261). It should be noted that Moore’s notion is so vague that it hardly resembles in any way the current conception of supervenience, which has proven its usefulness in various fields of philosophy and cultural studies. Moreover, Moore did not even use the term ”supervenience” in his writings.

It has been suggested that the term ”supervenient” might have entered our philosophical vocabulary from Latin translations of Aristotle’s *Nichomachean Ethics* (see Lewis 1985, p.159, footnote 4). Though, the credit for shaping the concept in its present form and introducing the term itself apparently belongs to R.M. Hare, who published his classic work *The Language of Morals* (1952) thirty years after Moore’s *Philosophical Studies*. Hare developed Moore’s idea a step further in the following, much cited passage, which is also thought to be the initial source of the current conception of supervenience:

> First, let us take that characteristic of ”good” which has been called its supervenience. Suppose that we say ”St. Francis was a good man”. It is logically impossible to say this and to maintain at the same time that there might have been another man placed exactly in the same circumstances as St. Francis, and who behaved in exactly the same way, but who differed from St. Francis in this respect only, that he was not a good man. (1952, p.145)

What Hare is doing here is that he is construing a relation between the valuational term ”good” and the terms denoting patterns of behavior or traits of character, which a ”good” person ought to possess. To put it more modernly, Hare is stating that supervenience is a dependency relation between valuational properties and descriptive or natural properties. Valuational properties supervene on natural properties in the sense that they are dependent on and totally determined by the latter: However, this does not mean that valuational properties could be solely defined *in terms* of natural properties, or *as* natural properties. In other words, if two persons share the same natural properties, they must also share the same valuational properties, and correspondingly, if two persons differ in respect of their valuational properties, they must also differ in respect of their natural properties.

(2.) The concept of supervenience was introduced to the mind-body debate by Donald Davidson in the early 1970s. According to him, mental characteristics are dependent or supervenient on physical characteristics, in the same way that Moore and Hare thought moral properties to be supervenient on natural properties (see Davidson 1970a;1973a). Davidson specifies this idea in ”Mental events” by stating that ”supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect” (1970a, p.214). Since Davidson did not endorse ontological reduction of mental events he had to come up with some conception which would not only justify antireductionism, but which would also make some sense to the token physicalist idea that every mental event is a physical event. Against this background supervenience seemed to be ideal for Davidson’s purposes. By stating that mental events supervene on physical events Davidson is able to complete AM and maintain its coherence without falling into dualistic ontology: mental events are supervenient features of physical events, which are strongly dependent on the physical domain without requiring the existence of psychophysical bridge laws.

In the current discussion of the mind-body relation supervenience is usually thought to hold between properties, which, indeed, was the original position of Moore and Hare. Hence Davidson’s talk of supervenience for events is, if not unusual, certainly rare. I will return to the issue of whether supervenience should be postulated for events, properties, or some other entities later, but from now on the concept will be interpreted as designating property supervenience. This will have no bearing on the following because what is of interest here is the general form of explanation.

(3.) In order to function as a useful concept Davidson’s notion of supervenience requires a few specifications. Kim has stated in ”Concepts of Supervenience” (1984a, p.156) that supervenience should not be seen to hold between predicates or properties but between *sets* or *families* of them. The idea is very commonsensical. For instance, the supervenient property ”to be a good person” might include every little detail (such as the person’s weight, height, date of birth, personal history, environmental effects etc.) among the set of base properties. Naturally, another person can be just as ”good” without having identical body and appearence, or sharing identical personal history with the first one. He might still be benevolent and honest and, hence, satisfy all the general criteria set for anyone to be called ”good”. Kim also adds that a supervenient property might have several alternative supervenience bases. For instance, one person might be humble and benevolent and the other one, in turn, courageous and honest. Both persons have individual properties that very well allow them to possess the supervenient property ”of being a good person”. Of course, they both would be ”good” in a different sense, because they have some ”good” properties but lack others.

With these examples, a distinction is generally made between *a minimal base* and *a maximal base*, or between *minimal base properties* and *maximal base properties* (ibid., p.165). A base that includes the properties of being courageous, honest, benevolent, humble etc. could be considered as maximal, because it suffices by itself to constitute every one of the properties of the supervenience family of ”being a good person”. The person that hads maximal base properties could be both courageous and humble but also honest and benevolent. A minimal base is, in turn, thought to be such that any property weaker than it cannot constitute the supervenience property

in question and hence cannot be a supervenience base at all. For instance, benevolence or honesty could be minimal properties, because anyone who possessed either of them is classified as ”a good person”. On the other hand, if someone does not have any of these minimal properties, he or she can in no way be said to possess the property of ”being a good person”.

(4.) The current discussion of supervenience distinguishes three distinct versions of supervenience: ”weak”, ”strong”, and ”global” supervenience. All three formulations are identical in respect to the concept itself, but each of them gives a different strength or a range of validity for supervenience relations. For instance, ”weak” supervenience can be defined as follows:

> [WS]A *weakly supervenes* on B if and only if necessarily for any x and y if x and y share all properties in B then x and y share all properties in A –that is, indiscernibility with respect to B entails indiscernibility with respect to A. (Kim 1984a, p.168)

The reason why this formulation of supervenience is called ”weak” is that it guarantees the dependency relation to hold only within any possible world, not among all possible worlds. To illustrate this point, let us say that A contains the supervenient mental property M, ”to feel pain in the right thumb”, and that B contains as the base property the neural state N, ”a certain excitation of the C-fibres which brings about feelings of pain in the right thumb”. According to weak supervenience, within any possible world there cannot be two persons, who share N but not M or *vice versa*. In other words, weak supervenience guarantees that when the base properties are set in any possible world, the supervenient properties must also be set in that same world: within any possible world there cannot be two entities agreeing in B but diverging in A.

However, this argument also contains the shortcomings of this definition. Weak supervenience allows there to be possible worlds within which everyone has the neural state N but not the mental state M. It also facilitates the existence of such possible worlds within which people feel pains in their right thumbs without ever having their C-fibres excited. Both of these options are totally legitimate because they fulfill the requirement set for weak supervenience; that is, that indiscernibility with respect to B entails indiscernibility with respect to A. This is an awkward situation. In fact, its problems are very much the same as those of emergence, which initially led to scholars abandoning the concept and turning to supervenience. This is why no one has seriously set out to defend any kind of weak supervenience; it is simply out of the question.

Since weak supervenience fails to capture the intuitive idea of dependence, which was expected of the concept, there have been efforts to strengthen the definition. For instance, Jaegwon Kim has presented the following definition that he calls ”strong” supervenience:

> [SS] Necessarily, for any object x and any property F in A, if x has F, then there exists a property G in B such that x has G, and *necessarily* if any has G, it has F. (1987, p.316)

Now, this formulation definitely circumvents all the problems related to the weak version. Strong supervenience guarantees that if any person or thing has the neural state N and the mental state M in any of the possible worlds, then every person or thing that has the same neural state N also has the mental state M, and that any person or

thing that has the mental state M also has the neural state N regardless of which of the possible worlds he or it might be in.

Many philosophers, Kim (1984a) and Beckerman (1992b) among them, have considered the strong version to be the only viable option as an explanation of the mind-body relation. This conclusion has been rejected at least by Grimes (1988). Furthermore, some scholars – such as Haugeland (1982) and Horgan (1982) – have gone to the defence of antireductionism and stated that both weak and strong versions are unacceptable because they concentrate too much on particular objects and on the relations of the properties these objects possess. They have developed an alternative view called "global supervenience", which is free of the commitment to property-to-property connections. The supervenience of the mental and the physical is, instead, based on the notion that worlds that are physically indiscernible are also psychologically indiscernible. This position can be defined as follows:

> [GS] For any worlds w_j and w_k, and for any objects x and y, if x has in w_j the same B-properties that y has in w_k, then x has in w_j the same A-properties that y has in w_k. (Kim 1987,p.317)

However, Kim has presented fairly convincing proof of the fact that global supervenience is also too weak to be accepted as an explanation of the mind-body relation. In order to refute global supervenience, one has to come up with a situation in which there are a pair of worlds that are physically indiscernible but psychologically discernible. In the following passages, taken from *Philosophy of Mind*, Kim gives an example of such a case. First he offers a starting point where

> There are two worlds, w_1 and w_2, and one mental property, *M*, and one physical property, *P*. There are two individuals *a* and *b* in both worlds. In w_1, *a* has *P* and *M*, and *b*, too, has *P* and *M*; in w_2, *a* has *P* but not *M*, and *b* does not have *P*. (1996, pp.225–226)

Now, Kim concludes that since worlds w_1 and w_2 are actually possible worlds, and that *a* and *b* are distinct individuals capable of existing without each other, it is justifiable to postulate the existence of the following two worlds, w_3 and w_4, within which

> *a* is the lone individual in w_3 and in w_4. In w_3 *a* has *P* and *M*; in w_4 *a* has *P* but not *M*. (ibid., p.226)

The worlds w_3 and w_4 are evidently physically indiscernible and yet psychologically discernible, because w_3 has *P* and *M* but w_4 has only *P*. On the basis of this, Kim concludes that global supervenience ought to be rejected, for it allows us to postulate pairs of worlds (such as w_1 and w_2) which entail the existence of another pairs of worlds (such as w_3 and w_4), which end up contradicting the indiscernibility principle. This suggests that the strong version is ultimately the most promising and unproblematic formulation of supervenience.

(5.) As I mentioned earlier, supervenience does not necessarily have to be thought to hold between properties. One of the reasons why the concept has such a wide range of applicability is that we can just as well speak of supervenience for facts, predicates, sentences, propositions, languages –or for events, as Davidson did. However, I am going to take supervenience here exclusively as a relation between properties. This is because I am inclined to

agree with Jaegwon Kim's opinion that, in case of the mind-body relation, discussion about events and predicates is unsuitable. Kim's view is dictated by two beliefs: (i) that strong supervenience (which we just established as the only viable version of the concept) is so forceful that it always implies the existence of nomological biconditionals and hence reduction, and (ii) that property supervenience is fundamental. (see e.g. Kim 1984a)

Kim's idea that strong supervenience implies reductionism can be explicated in the following way (see Beckerman 1992b, pp. 97–98) . Say, we have a set of subvenient properties B which contains three basic properties G_1, G_2, and G_3 and all the combinations that are constructable from them by such methods as conjuction, complementation and so on. Let us also assume that an object *a* has the properties G_1 and G_2 but that it lacks the property G_3. Let us say that B_1 is a maximal property which determines which subvenient properties an object has and define B_1 by "*x* has B_1 iff x has G_1 and G_2 and $-G_3$". We may also add that *F* is a member of family A of properties which strongly supervene on B. Since *a* has the properties G_1, G_2 and $-G_3$, it evidently has B_1, and since it has B_1, it also has the supervenient property *F*. This dependency can be expressed by the generalization

For all *x*: if *x* has B_1, then *x* has *F*

which holds with necessity. In other words, any *x* object that has B_1 is B-indiscernible and hence has also the same A-properties as *a*, which in this case means the property *F*. Clearly B_1 is not the only B-maximal property, for there are eight different variations of the form "*x* has G_1 and $-G_2$ and $-G_3$" etc. Some of these basic properties suffice to constitute the supervenient property *F*, but some (such as "*x* has $-G_1$ and $-G_2$ and $-G_3$") do not for obvious reasons. If, say, B_1, B_3, and B_8 do suffice to constitute the A-property *F* and B_2, B_4, B_5, B_6, and B_7 do not, we may define the property B_F by "*x* has B_F iff *x* has B_1, B_3 or B_8" and the property B_{-F} by "*x* has B_{-F} iff *x* has either B_2, B_4, B_5,B_6 or B_7". Therefore our definitions provide us with two nomological biconditionals:

For all *x*: *x* has B_F iff *x* has *F*, and
For all *x*: *x* has B_{-F} iff *x* has *-F*.

Kim interprets that the existence of these biconditionals opens up the possibility that properties *F* and *-F* are reducible to the basal structure, in other words to certain B-properties.

Kim's belief that property supervenience is fundamental is closely linked with the idea that supervenient properties are reducible to their bases. This link is described in the following passage:

> First, we need to be sensitive to the distinction between *predicates* and *properties*, and beware that complexity or artificiality attaching to predicates (or linguistic constructions in general) need not attach to the properties they express. A long Boolean combination of predicates would normally be complex *qua* predicate; on the other hand, the property it expresses need not inherit that complexity...--- When we speak of laws, we may have in mind either sentences or some nonlinguistic, nonconceptual, objective connections between properties. If laws are taken to be sentences, our results do not show that psychophysical supervenience entails the existence of biconditional laws. For we are given no guarantee that there are predicates, especially reasonably simple and perspicuous ones, to represent the constructed properties. Reformulating our basic definitions in terms of *predicates* rather than properties will not help; for that would make infinitary procedures highly dubious, perhaps unacceptable. Moreover, strong psychophysical supervenience stated for psychological and physical predicates seems considerably less plausible than when stated

> for properties...What is the physical predicate that entails, say, ”being bored”? It seems that we would at least need to appeal to ”ideal physical languages” and the like to get started, and this might bring us right back to talk of properties. (1984a, p.172)

In a nutshell, Kim believes that supervenience for most entities can be explained in terms of property supervenience, and that it should be done in this way. The motivation behind this claim is the notion that strong supervenience always entails some sort of psychophysical biconditionals. Since subvenients and supervenients are too complex to be expressed in predicates, and since by using predicates we might lose psychophysical biconditionals, psychophysical correlations ought therefore to be expressed in terms of property supervenience.

Kim’s claim is undeniably coloured by his desire to find a way to facilitate mind-body reduction. Hence it is only natural that those in favour of antireductionism have reacted against Kim’s views. For instance, Jerry Fodor (1974) and Paul Teller (1984) have stated that even if one accepts that there are psychophysical biconditionals and that supervenience holds between properties, this does not necessarily mean that supervenient properties are reducible to physical properties. Fodor and Teller have set an additional criterion for reduction according to which the base properties have to be instantiated by events or objects that fall under the laws of physics. Ansagar Beckerman (1992b, p.116) has, in turn, noted that reduction might be carried out even without correlating bridge laws because mental states might still be microreducible. Or, correspondingly, the reduction might equally well fail even in the presence of bridge laws, if the bridge principles themselves are ”unique”; that is, if they can be discovered only by studying the very same processes they are supposed to be governing.

However, more severe problems with Kim’s view lie elsewhere. At one point, Kim writes that ”such operations as infinite conjunctions and infinite disjunctions would be highly questionable for predicates, but not necessarily for properties – any more than infinite unions and intersections are for classes”(1984a, p.172). Surprisingly, Kim seems to support the view which accepts that supervenient properties (even mental ones) can be realized by disjunctive or conjunctive physical properties. I listed earlier a number of arguments against disjunctive and conjunctive properties, which rejected this claim. Luckily, Kim has since then re-evaluated his position and stated that

> ...given the extreme heterogeneity of the actual and possible physical realizers of [any mental state]...we cannot expect [a disjunctive property] to be a well-behaved property that captures a point of significant resemblance among items that have it. (1996, p.219)

So, it seems that Kim does not approve disjunctional properties after all. This also means that he still has cope with all the problems and obstacles set by the multiple realization of mental states. Before this challenge is met, all the efforts Kim makes to defend property supervenience are futile. Property-to-property reductions are particulary vulnerable to MR-based counterarguments, since the evidence against them is so overwhelming.

(6.) To summarize our discussions of supervenience, it might be said that the concept of supervenience does strengthen the link between the mental and the physical more than the concept of emergence. Whether it manages to carry out the ”dependency without reduction” idea any better is questionable. It became very clear that the strong version especially tends to insist on over-tight nomological connections, strong enough to satisfy psychophysical reductions – at least in principle. Because of this defect, supervenience does not offer final solution

to the mind-body problem. In fact, it has only given a new framework to the seemingly endless debate over the reductionism/antireductionism issue. There are just as many arguments for both views as there were before the application of the concept.

4.3 Obstacles to Mind-Body Reduction

Now that the central issues related to the concept of reduction and to the mind-body relation have been clarified we are in a position to summarize most of the general views taken on the mind-body relation. We have learned that both the concept of reduction and the mind-body relation can be understood in ways which by themselves exclude the possibility of a mind-body reduction. First of all, reduction can be seen to hold between two scientific theories where one is derivable from the other, although, there is strong evidence supporting the notion that this approach cannot be successfully carried out. Reduction can, however, also be postulated to hold between properties, events, or states. We have seen that this latter approach, in turn, has at least some convincing theoretical justifications. Secondly, mental states can be given such a strong ontological status that they cannot be reduced even in principle. On the other hand, stronger versions of supervenience seem to actually imply mental reduction. These are only the extreme positions of the debate, and there are naturally numerous intermediary cases. In the following, I will try to give a balanced picture of the debate and highlight, once again, some of the most important views held on mind-body reduction.

Let us start with the case of a Nagelian inter-theoretic mind-body reduction. There are a number of substantial reasons why the idea that a psychological theory is reducible to some physical theory should not be accepted. For instance, Hilary Putnam (1975) has stated that folk psychology is autonomic with respect to neurobiological theories and hence irreducible to them, because its logical relations cannot be reduced to causal relations. But there are also other approaches. To begin with, it is generally agreed that there is no complete theory of, for instance, folk psychology, which could function as the target theory of the reduction. In fact, this incompleteness of folk psychology has been one of the motives behind eliminative materialism: since the nature of folk psychology itself enschews inter-theoretic reduction, it should be judged as a ”false” theory and abandoned as a whole.

However, the case is not so simple. Many philosophers, T. Hogan and J. Woodward (1985) among them, have defended the status of folk psychology against these claims, and stressed that it is still a vital theory. Daniel C. Dennet (1987) has, in turn, stated that folk psychology is a useful tool in prediction and interpretation of mental states and events, when it is understood as a rational calculus. Moreover, the rejection might be premature, since a comprehensive and unanimous formulation of folk psychology does not even exist. The latest interpretations of folk psychology as ”mental simulation”, presented by Robert M. Gordon (1986) and Alvin I. Godman (1986), show that the discussion is still unfinished.

Furthermore, this issue also has another side to it. The ”elimination” of folk psychology has the ulterior motive that psychological theory is to be replaced with a neuro-computational one. We may just as well ask whether there is any such comprehensive theory available. Yet again the answer is negative. In addition, many cognitive psychologists, such as Jerry A. Fodor (1975), have doubted that the neural sciences could even provide

an adequate amount of information needed to formulate such a physical theory – at least at their present stage of development.

There is also a second line of argument against the mind-body reduction, which governs inter-theoretic reduction but is not necessarily bound to it; that is, the absence of bridge laws. The Nagelian derivation requires the existence of appropriate bridge principles in order to be feasible, but so too do local state-to-state reductions. Probably the most forceful argument against reduction in this case comes in the form of the Multiple Realization Thesis. The fact that mental states are multiply realizable refutes any kind of psychoneural identification and also the existence of bridge laws. As stated earlier, this is mainly due to the fact that disjunctive properties are unacceptable. Since there cannot be property-to-property correlations but only property-to-disjunctive property connections (and the latter are not valid) there can be no psychophysical bridge laws.

Additional counterarguments are provided by those in favour of emergence. The notion initially put forward by Alexander and developed in the work of Popper claims that emergent properties are novel in respect of the lower level and that they could also have new causal powers. The more radical emergentists have interpreted this to mean that to be emergent is equivalent to having new causal powers. In case of the mind-body relation, this means that the essence of being conscious is to have the ability to causally interact with the physical domain. Proponents of this view think that mental causation (and hence consciousness) would disappear in reduction, since all mental actions become solely explainable in terms of physical processes and events. The proposers of the ontological version of emergence – such as Popper (1977), Mario Bunge (1977), and Joseph Margolis (1978) – have thus been particularly enthusiastic about denying the existence of bridge laws.

We should also keep in mind Donald Davidson (1970) and his Anomalous Monism, which also denies the existence of strict psychophysical laws. Davidson's counterargument is also ontological in nature, since it rejects connecting, lawful psychophysical principles (although not every kind, only the strong ones) on the basis of token physicalism: there is no place for such principles, because physical and mental events are ultimately one and the same thing.

The third basic line of thought against reductionism is to take refuge behind the qualia problem. For instance, Saul Kripke (1972), Thomas Nagel (1974), Frank Jackson (1986), and John Searle (1992) have stated that the sensing of qualia (e.g. the smell of a rose or the quality of a headache) is a constitutive feature of the mental, which simply cannot be found from or explained by the physical processes. Naturally, this prohibits mind-body reductions in principle, which means that the position is not affected by the possible scenario according to which we may some day have perfect property-to-property connections with exceptionally strongly binding bridge laws.

I will summarise this discussion by concluding that the issue of mind-body reduction is not an easy one to tackle. Ansagar Beckerman's (1992b) discouraging observation, that the existence of psychophysical bridge laws is neither a necessary nor a sufficient condition for a mind-body reduction, is enough on its own to demonstrate the complexity of the matter. In reality, the issue of mind-body reduction is not about one specific problem or obstacle; it consists of several rather independent, occasionally overlapping discussions any one of which is enough to shake the sanity of any philosopher. It is clear that if Jaegwon Kim's reductionist theory of the mind proves to be a credible one, he must offer some surprising new insights.

5. Kim's View of Physical Realization

Now we can finally shift our focus to Jaegwon Kim's philosophy of the mind and especially to his type-identically reductive conception of physical realization. We have now glanced at all the major positions taken on the modern mind-body debate, and there has certainly been a considerable amount of ground to cover. However, I believe that the background work has proven to be worth the effort. We have so far learned that antireductionist versions of functionalism are the most prominent theories to date. Faith in functionalism has occasionally appeared so strong that it has frequently been considered as the final solution to the mind-body problem. Donald Davidson's Anomalous Monism too has won many scholars over as the only viable antireductionist alternative to functionalism.

The reason why I wish to bring Kim's view into this discussion is not that he offers a theory superior to any of the aforementioned positions. In fact, I do not think that Kim's conception is, in the end, any more flawless or plausible than the theories it is criticizing. Nevertheless, Kim manages to point out such crucial defects in the pre-eminent positions taken regarding the philosophy of the mind that they simply cannot be passed over. Moreover, Kim's manner of taking the implications of his findings seriously and developing them to their radical, reductionist conclusions is rather appealing in its freshness. Only these sorts of fearless attempts have any real chance of breaking away from the circle within which the mind-body debate seems to have been stuck for too long a time.

Kim's own theory of the mind, which I am about to present here, is guided by two beliefs; that (i) all versions of emergence and nonreductive physicalism are committed to downward causation, and that (ii) this commitment is bound to lead them either into unacceptable epiphenomenalist formulations or into a more preferable alternative, which is reductionism. I will start with the arguments and assumptions that Kim has formulated in support of these two claims and then proceed to consider the type-identical reductions of mental states, which Kim believes to result from the claims.

5.1 Nonreductionist Physicalism as Emergence

All the current theories of the mind seem to have absorbed the idea of the world as a layered structure. On the most elementary level this view rejects Cartesian substance dualism but it also endorses a fairly advanced ontological picture that has its roots in emergence. The world is seen to consist of an array of levels, within which each level of entities has its characteristic and distinct set of properties. The hierarchy is thought to start from an ultimate base level containing the basic particles of micro-physics (such as electrons, neutrons etc.), all characterizable by basic fundamental physical properties and relations (e.g. mass, energy, and spin). The new, ascending levels are structures created from the entities belonging to the lower levels. To be more precise, the levels are constructed from two components; that is, a set of entities constituting the domain of particulars for each level and a set of properties defined over each domain. Furthermore, relations between distinct levels are thought to be *mereological* in nature, which means that the higher levels are mereological sums or compositions

of the lower level entities. Naturally, this has the consequence that particles of distinct levels are seen as members of a part–whole relationship.

Now, this last issue brings us right back to the mind-body problem: If entities at distinct levels are ordered by the part-whole relation, what sort of a relation holds between properties associated with these levels? How are mental properties related to physical properties? Almost everything I have so far written is somehow related to this question, and it has become apparent that there are numerous different and even mutually exclusive answers. However, Kim has taken an interesting approach to this issue which is able to bring all the prominent alternatives into the same category. As was noted, in the contemporary debate on the mind-body problem the prefered theories are either versions of nonreductive physicalism or emergentism. In addition, it was made clear that the mind-body relation itself can be successfully explained only in terms of strong supervenience. We have also learned that the concept of physical realization (which has not yet been properly explicated) is another fashionable notion that is often used in connection with supervenience to characterize the relation between mental and physical properties. Kim presents two claims concerning this general picture; that a) nonreductive physicalism comes so close to emergentism that it could and should be considered as a version of emergentism, and that b) the concept of physical realization entails supervenience.

Let us start with the latter claim regarding the converse entailment of supervenience and physical realization. It has already been mentioned that the idea of physical realizability originates from Putnam's Multiple Realizability Thesis which he presented in the 1960s, and which resulted in the disappearance of all reductionist formulations. Kim has himself explicated the concept of physical realization as follows:

> To say that a physical state, P, ”realizes” (or ”instantiates” or ”implements”) a mental state, M, in an organism or structure of kind S must at least include the claim that in that kind of organism M occurs at a time, as a matter of law, just in case P occurs at that time. The claim that the excitation of C- and A-delta fibres ”realizes” pain in humans is, or must include, the claim that humans are in pain just in case their C- fibres or A-delta fibres are activated. Thus, corresponding to the thesis of property emergence is the nonreductive physicalist's claim that *mental properties are instantiated only by being realized by physical properties in physical systems*. (1992a, pp.131–132)

We can clearly see that Kim's formulation is in agreement with the general notion of physical realization which is explicated, for instance, by LePore and Loewer in the following way:

> Exactly what is it for one of an event's properties to *realize* another? The usual conception is that *e*'s being *P* realizes *e*'s being *F* iff *e* is *P* and *e* is *F* and there is a strong connection of some sort between *P* and *F*. (1989, p.179)

However, besides the fact that LePore and Loewer have formulated their conception primarily for events, there is another major difference with respect to Kim's view. Lepore and Loewer see physical realization above all as an explanatory, epistemic relation, while Kim stresses that the relation between, say, a physical property *P* and a mental property *M* is an objective and metaphysical one. This becomes apparent in the quotation taken from ”Non-reductivist's troubles with mental causation”:

> ...I am taking a realist attitude about explanation: if *P* explains *M*, that is so because some objective metaphysical relation holds between *P* and *M*. That *P* explains *M* cannot be a brute, fundamental fact about *P* and *M*...In the case of realization, the key concepts, I suggest, are those of

> ”causal mechanism” and ”microstructure”. When *P* is said to ”realize” *M* in system *s*, *P* must specify a micro-structural property of *s* that provides a causal mechanism for the implementation of *M* in *s*...in fact, if we are speaking meaningfully of ”implementation” of *M* –*P* will be a member of a family of physical properties forming a network of nomologically connected micro-structural states that provides a micro-causal mechanism, in systems appropriately like *s*, for the nomological connections among a broad system of mental properties of which *M* is an element. (1993b, p.197)

We will soon find out that a metaphysical interpretation functions as the basis for Kim’s reductionist views.

”Physical realization” is evidently a concept that is used to explain psychophysical property relationships. It basically states that psychological and neurophysiological properties are related such in a way that the latter realises the former. Moreover, the concept requires ”a strong connection of some sort” (to use LePore and Loewer’s terms) between mental and physical properties. In fact, the concept requires that *P* __*M* holds with *nomological necessity*. The reason for this is that physical realization is supposed to explain the psychophysical realization, and this cannot be achieved through only physically necessary connections – or at least through each one of them (LePore & Loewer 1989, p.179). On the basis of this, Kim concludes that physical realization does entail stronger versions of supervenience, because the conditions of explanation and nomological necessity cannot otherwise be met. Though, Kim acknowledges that this entailment does not hold in every case, especially if weaker versions of supervenience are applied or if the concept has not been used at all (1993b, p.196). However, we have seen fairly convincing proof of the fact that strong supervenience is the only viable option when it comes to explaining the mind-body relation.

Let us turn to Kim’s other claim, which states that nonreductive physicalism is so close to emergentism that it can be considered as a form of emergence. According to Kim, the similarity of these two views is based on the following four principles or features, which they share with each other (see Kim 1993b, pp.198–201)

> 1. (*Physical monism*) All concrete particulars are physical.
>
> 2. (*Antireductionism*) Mental properties are not reducible to physical properties.
>
> 3. (*The Physical Realization Thesis*) All mental properties are physically realized; that is, whenever an organism, or system, instantiates a mental property *M*, it has some physical property *P* such that *P* realizes *M* in organisms of its kind.
>
> 4. (*Mental Realism*) Mental properties are real properties of objects and events; they are not merely useful aids in making predictions or fictitious manners of speech.

About the first two features nothing further need be said. I have previously explained in detail that the concepts of both emergentism and nonreductive physicalism accept only a purely materialistic ontology of concrete physical objects and events. Furthermore the reasons why mental states are thought to be irreducible to physical states has been thoroughly explained.

The third feature, however, might require some thought. The idea of physical realization endorsed by nonreductive physicalists states that whenever a mental property is instantiated in a system, the case is such that the system has instantiated an appropriate physical property, which in turn has instantiated the mental property. Furthermore, the instantiation of the mental property must follow the instantiation of the physical property as a

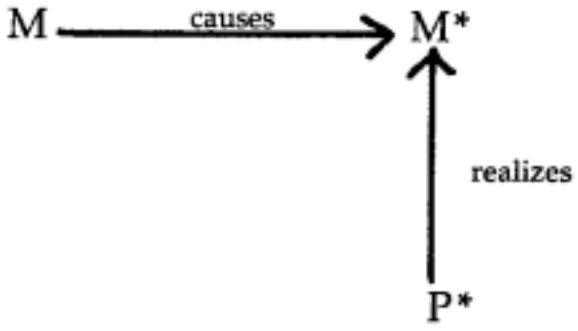
matter of necessity. The same notion can be found in emergentism. As explained earlier, emergentists believe that, whenever the appropriate basal conditions for the appearence of higher-level properties were present, those properties had to necessarily emerge.

Finally, the fourth feature is implicitly contained in the above three principles. The notion of mental realism was particularly strongly present in the works of emergent evolutionists; they thought that the world had reached its present form in the course of emergent evolution, which had brought about true complexity and fullness. Realism with regard to mental states is also apparent in the works of nonreductive physicalists, who have constantly defended the autonomy of the mental and stressed the special status of psychology as a scientific practice.

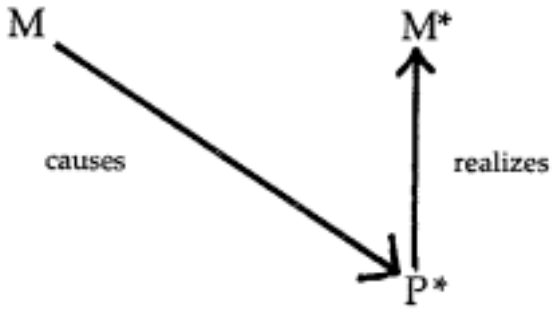
5.2 The Problem of ”Downward Causation” and the Charge for Epiphenomenalism

Kim combines the claims, that nonreductive physicalism is a form of emergence and that physical realization entails strong supervenience, as an argument against nonreductive physicalism. Kim states that the two claims force nonreductive physicalism to accept ”downward causation”, which is, in itself, an unacceptable conception. Kim starts with the idea that nonreductive physicalism requires the existence of mental causation. In case of emergence, this is indisputable. One essential feature of emergence is that emergent levels have novel and irreducible causal powers with respect to the lower levels. In fact, the stronger versions of emergence, such as the one proposed by Samuel Alexander, straightforwardly stated that the possession of these sorts of causal powers was a necessary condition for being an emergent property. Nonreductive physicalism has also features that require consciousness to possess novel and irreducible causal powers. The idea of the autonomy of the mental and the idea of psychology as a special science seem to suggest that psychology governs causal connections that cannot be captured by the underlying sciences (Kim 1993b, p.204). Why else should the mental domain be considered as autonomous or psychology be taken as an autonomous science?

But what is so wrong with the suggestion that the mental might have irreducible causal powers? Or what is the matter with the concept of downward causation? Kim answers these questions with a simple but powerful argument. If any mental property has causal powers, it must be able to manifest these powers by being causally efficacious with respect to another property, mental or physical. Let us say, that a mental property *M* causes another mental property *M** to be instantiated on a certain occasion:

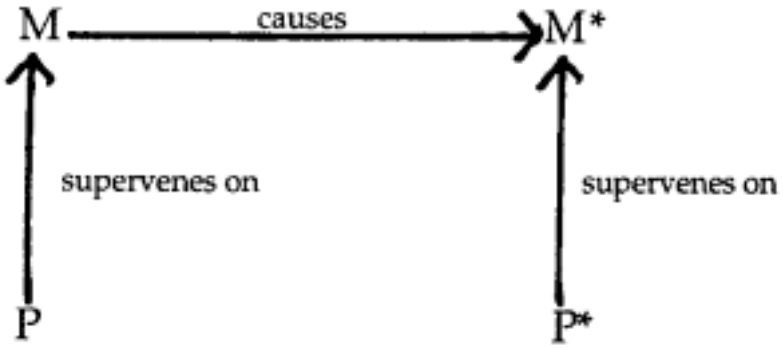


Naturally, the mental property *M* is a physical realization of a physical property *P*. Correspondingly, the mental property *M** needs a physical realization base in order to be instantiated. The mental property *M** is there only because it is realized by a physical property *P**; in fact, *M** cannot even exist without *P**. Now the situation is such that, in order for *M* to cause the instantiation of *M**, *M* has to cause the instantiation of the physical realizer of *M**, the physical property *P** :



The case, where *M* brings about *P** through mental-to-physical causation, is obviously an example of downward causation. (Kim 1993b, p.205; for a repetition of the argument in case of emergence, see Kim 1992a, p.136)

There are only two problems with this notion. Firstly, *P** is solely sufficient for bringing about the instantiation of *M**, which seems to pre-empt *M*'s role as the cause of the realization of *M**:



Of course, someone might argue that *M* brings about *M** by causing the occurrence of *P**. This would require acceptance of the Principle of Causal Individuation of Kinds, which states that objects and events fall under kinds insofar as they have similar causal powers (see Kim 1992b). The idea would then be that *M* had the same causal powers as its physical realizer *P*. However, even this view leads to an unacceptable situation, because both *M* and *P* would be sufficient to cause *M*; M** would have two distinct *causes* and be thus *causally overdetermined*. If this conclusion is to be avoided, mental states cannot be considered as scientific kinds. This has the consequence that mental states supervene on physical states and are realized by them but that mental states are not *caused* by physical states. The only way in which a nonreductionist can assign causal powers to mental states is in terms of the Causal Realization Principle, which states that if the occurence of *M* is realized by *P*, then any cause of *M* must be the cause of the instance of *P* (see Kim 1993b, p.205). The idea is that mental states do not have

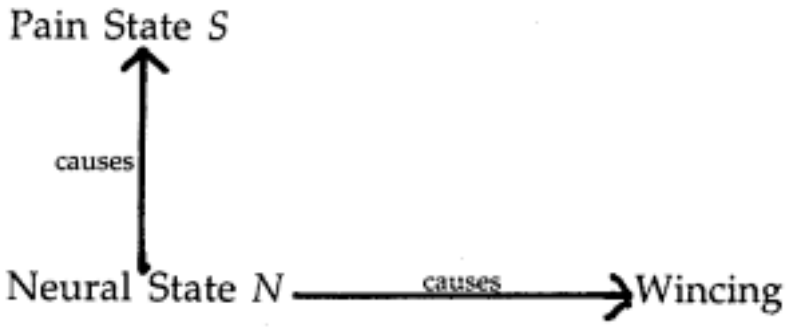
independent causal powers, but that they inherit them from their physical realizers. Because mental states are not genuinely causally efficacious causal overdetermination can be avoided.

Secondly, physical realization entails strong supervenience, which means that the mental property M is dependent on and determined by the physical property *P*. Since the only way *M* can cause *M** is to bring about *P**, *M* has to somehow transform *P* into *P**. This last manouver is, however, totally impossible. In the light of the strong dependence provided by the supervenience relation, it is not even remotely possible for *M* to change the very same physical realization base *P* by which it is totally determined.

Kim believes that the argument above demonstrates conclusively that nonreductive physicalism always implies downward causation, and that downward causation is an intolerable conception which must be abandoned. However, if the possibility of downward causation is excluded, mental states have no means of being causally efficacious any longer. Non-reductive materialism is thus committed to epiphenomenalism.

But is there something wrong with being an epiphenomenalist? In fact, there is one good reason not to endorse this view. I am going to demonstrate it by taking Davidson's Anomalous Monism as an example. I chose to consider Davidson's view mainly because it has been accused of being a form of epiphenomenalism by a number of philosophers – incuding Honderich (1982), Kim (1984c; see also 1993a), Sosa (1984), Stoutland (1985), and Fodor (1989).

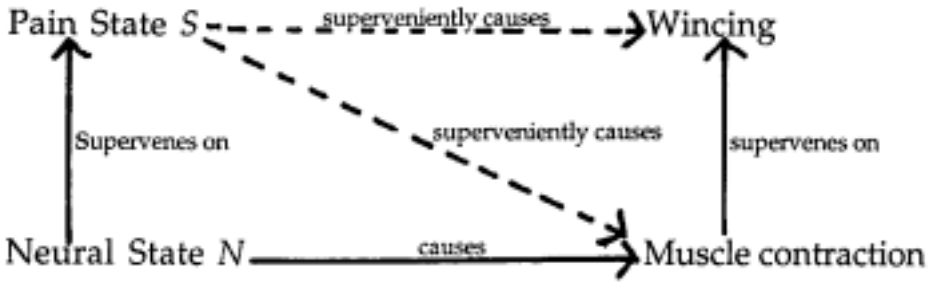
Let us consider a situation in which the occurrence of a certain neural state *N* is found to be correlated with a certain pain state *S*. Let us also assume that the occurrence of the pain state *S* is followed by winces. According to Anomalous Monism, the neural state *N* and the pain state *S* are only two tokens or aspects of the same event. Furthermore, their relation is thought to be such that *S* supervenes on *N*. Naturally, it would also be assumed that the occurrence of pain caused wincing. However, if Kim's claim about the impossibility of downward causation is correct, there is no way wincing could be caused by the occurrence of the pain state *S*. Instead, it is caused by the neural state *N*. An epiphenomenalist model of the situation might look something like the following:



The pain state *S* is seen only as a by-product of the neural state *N*. Neither has it any causal powers left, for all the causal work is done at the physical level by the neural state *N*. What is notable about this model is that the pain state *S* is no longer thought to be supervenient on the neural state *N* but to be *caused* by it. This is due to one of the central theses of epiphenomenalism which holds that mental states are always *causally produced* by physical states (see e.g. Shaffer 1968, pp. 68–69). This is the reason why epiphenomenalism is to be rejected. It is very

difficult to think that a physical state could have a causal relation with the very same mental state it instantiates. Futhermore, it is impossible to imagine the specific nature of the relation in question.

Since epiphenomenalism is an unacceptable view, an alternative model is needed. Kim (1996, pp. 150–151) has suggested that such a model might be formulated in terms of supervenient causation. The idea is that mental causation is supervenient on physical causation in the same way as mental states are supervenient on physical states. For instance, if a person suddenly feels pain and winces, it could be said that the pain ”superveniently causes” wincing. The idea is that the physical level causal relation between the neural state *N* (which physically realizes the pain state *S*) and a muscle contraction (which physically realizes wincing) constitutes a supervenience base for a supervenient causal relation between the pain state *S* and wincing:



Kim illustrates the idea by giving an analogy in which the heating of a kettle causes water to boil. The only truely existing phenomenon in this example is the physical, microcausal relation between the increased kinetic energy of water molecules and violent ejections of water molecules into the air. However, in our ordinary language we usually refer to the macrophenomenon, which is the boiling of water that is thought to be caused by the heating of the kettle. Naturally, this latter phenomenon is supervenient on the microphenomenon.

Even though the notion of supervenient causation is very commonsensical, its ontological status is a something of a mystery. What is actually meant by the term”superveniently causal”? I believe the best way to interpret the term is to regard it merely as a figure of speech which is commonly used in our ordinary language. Hence supervenient causation is merely a cultural entity that exists relative to our minds. As long as people find it easier to talk about the boiling of water or the fact that pain causes wincing, instead of using the terms of chemistry or physics, superveniently causal phenomena will remain. Nevertheless, it does not change the fact that there is no boiling of water or mental causation – just molecular and neural processes.

5.3 Type-identical Reductions of Mental States

The model of supervenient causation is without a doubt a dream come true for a nonreductive physicalist. It manages to circumvent all the major pitfalls that are usually thought to undermine nonreductionism. Firstly, supervenience provides psychophysical property relations with a strong nomological dependence and a high degree of explanatory power without endorsing reduction. This is an essential goal that all the former versions of emergence failed to reach. Secondly, the model succeeds in maintaining the mental as causally efficacious by

reformulating mental-to-mental and mental-to-physical causations as forms of supervenient causation. Finally, by preserving mental causation the model is also able to protect the status of psychology as a special science; there seems to be causal relations that, after all, cannot be encapsulated by underlying theories but only by psychological theories. These facts strongly indicate that nonreductive physicalism is the way forward, and that it is time to put reductive views aside for good.

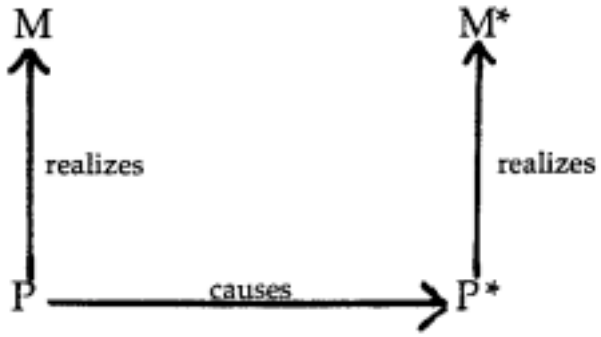
If the model of supervenient causation is the huge success as it appears to be, then why not settle for it? I can think of only one reason, which is an old one, but which is nonetheless valid; that is, the principle of ontological simplicity. In other words, we must ask if there are any reasons why Ocham's razor should not be used to dispose of supervenient mental properties and supervenient causation. When the question is put this way, I can suggest a number of reasons.

I would like to start with fact that the model of supervenient causation preserves mental causation only by re-interpreting falsely the Principle of Causal Inheritance, which initially denied the causal efficaciousnss of the mental. As was mentioned earlier, this principle states that mental properties inherit their causal powers from physical properties. Since the causal powers of the mental are derivable from physical causation, this evidently pre-empts any possibility of mental causation – at least in any meaningful sense. However, the model of supervenient causation takes this to mean that if the causal powers of the mental and the physical are identical, then any mental property *M* realized by some physical property *P* can do the same causal work as *P*. But what are we actually saying here? I tend to agree with Kim that this situation is parallel with a situation in which we say "the increase in the water temperature caused the boiling of the water", although what is really meant is something like "the moving of H_2O molecules with increasing velocities caused ejections of H_2O molecules into the air" (see Kim 1996, p.151). The fact that we speak in our ordinary lives about the boiling of the water does not change the real state of affairs according to which what is actually happening are occurrences on the molecular level. Correspondingly, it is only natural that we say that "Jones got angry because he accidently got a painful cut in his hand", when such a thing actually occurs. Instead, if I state in this situation that "Jones' such and such nerves are stimulated in such and such ways that it has caused such and such an occurence in his brain", people could be forgiven for being confused. Nevertheless, the situation would be identical in both cases.

What we are learning here is that there is no supervenient mental-to-mental causation starting from Jones' pain state and ending up in Jones' other mental state of being angry. Neither is there any kind of supervenient causal relation between Jones' pain state and that brain process which realized Jones' mental state of being angry. There is only a causal chain of physical processes starting from the nerve ends in Jones' hand and ending up in those parts of Jones' brain that process this information. Everything else is, literally speaking, semantics; that is, figures of everyday speech, metaphors, descriptions, and explanations that have no ontological significance. Of course, these expressions specifically designating either supervenient properties (e.g., "the boiling of water") or supervenient causation (e.g., "a steel rod lengthened, when it was heated") have their function in our ordinary, day to day communications, but they should not be allowed to enter the area of science or appear in scientific vocabularies.

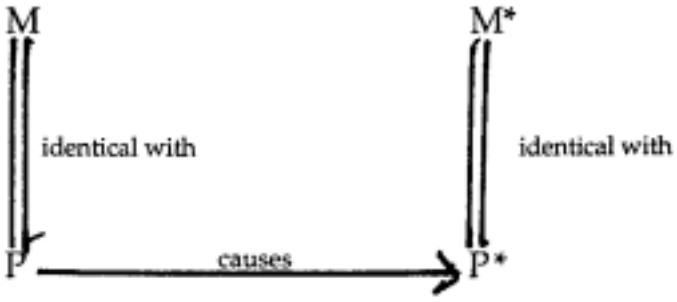
If we abandon the model of supervenient causation (as we should), what is left is an ordinary supervenience relation between mental and physical properties. However, I presented earlier Kim's claim that supervenience and

physical realization conversely entail each other: they are essentially identical in philosophical import. This gives us the chance to convert supervenience relation into the following physical realization model:



As the model illustrates, the only true causal relation that is left is the one that holds between physical properties *P* and *P**. The causal powers of *M* are derived from the causal powers of its physical realizer *P* the instantiation of which, in turn, totally determines the instantiation of *M*. This has the consequence that the instantiation of *M* causes whatever its physical realizer *P* causes, and whatever causes *P* to be instantiated also causes *M* to be instantiated.

According to the concept of physical realization, *P* is by itself sufficient to cause the instantiation of *M* and *P** is sufficient to cause the instantiation of *P**. However, since physical realization always entails strong supervenience, the conditionals "*P* _ *M*" and "*P** _ *M**" hold with nomological necessity. In fact, Kim has stated in his *Philosophy of Mind* (1996, pp. 233–234) that strong supervenience implies such a strong nomological correlation that "*P* _ *M*" and "*P** _ *M**" can be interpreted as biconditional bridge laws of the form "*P* __*M*" and "*P** _ *M**". The existence of appropriate bridge laws in turn facilitate the formulation of type-identical reductions of the form "*P* = *M*" and "*P** = *M**":



Naturally, this model is unacceptable in its present form. We have only to think of the Multiple Realization Thesis in order to perceive the crucial defect it has. We may thus very well postulate that, for instance, the mental property *M* has several physical realization bases, P_h in humans, P_m in Martians, P_c in chimpanzees and so on. Since disjunctive properties of the form P_1 v P_2 v, ..., P_n are not allowed, we have to conclude that Kim's model is not a credible one. However, Kim has acknowledged this problem and added an additional condition which

states that bridge laws are ”structure-restricted”. This means that we can find different psychophysical property correlations in a different organism, which hold only in the domain of that specific organism or structure. For instance, if *P* is a realizer of *M* in an organism or a structure of type *S*, which could in this case be the human brain, the following relationship holds as a matter of law (1996, p.233):

$$S \_ \ (M \_ P)$$

In other words, if we find through empirical research that *P* is correlated with *M* in the human brain, we may conclude that this is so in all humans (where they are healthy and normal). However, *P* might still be found to coexist with mental properties other than *M* in other organisms and *vice versa*.

Finally, I wish to stress that Kim does not believe that the reductionist model would facilitate the reduction of psychology in its entirety. It merely states that the neural sciences provides us with an increasing amount of information, on the basis of which it is possible to learn that some mental states are often correlated with certain neurophysiological processes. This in turn gives us a chance to carry out local, property-to-property reductions of mental states to their neurobiological bases. However, Kim (1996, p.235) does believe that even local reductions are enough to discredit the validity of psychological theories, or at least he implies that the range of their application might not be as broad as it is usually thought to be. After all, the reductionist model clearly undermines psychological properties as scientific kinds.

6. Further Considerations

This chapter has been dedicated to exploring some of the major views regarding the mind, which have been presented by a variety of philosophers during the latter half of this century. We have learned that since the 1960s different identity theories have received very little support, and that their proponents have become a minor group (although a persistent one) in the mind-body debate. The reason for the rise of antireductionism was argued to be Hilary Putnam’s Multiple Realization Thesis, which seems to entirely rule out the possibility of a mind-body reduction. It was also shown that Putnam’s MR launched a program of functionalism, which is still the prominent view on the mind and which has not yet lost its early promise. Only Donald Davidson’s peculiar Anomalous Monism has been able to challenge the popularity of functionalism.

It was found that the current positions of the philosophy of the mind are based on two inter-related notions; that is, on the concepts of supervenience and physical realization. It was claimed that philosophers have been inclined to characterize the mind-body relation in terms of supervenience because of the failure of the concept of emergence to provide a strong enough correlation between physical and mental states. The application of the concept of physical realization was in turn argued to be initiated by the central thesis of functionalism according to which mental states can be seen as mere abstract task-descriptions implemented or ”physically realized” by a variety of organic (or possibly even inorganic) systems.

During the last decade reductionism has started to regain support. This, it was argued, was for three reasons: Firstly, supervenience provides such a strong physical determination of mental states that it is in danger of losing mental causation altogether and leaving space only for epiphenomenalism. Secondly, there have been suggestions that only the ”strong” version of supervenience can guarantee a tight enough correlation for the mind-body relation. However, it has been claimed that strong supervenience is too strong in the sense that the logical necessity it implies allows the possibility of mind-body reductions. Thirdly, since supervenience is merely a logical relation, it cannot give any real explanation of the relation between consciousness and the brain. These three points, combined with dramatic improvements in brain research techniques, have led to the resurrection of reductionism.

I took the side of Jaegwon Kim and supported the view that strong supervenience is the only viable option as an explanation of the mind-body relation. I also agreed with Kim that strong supervenience allows local, type-identical reductions. My conclusion was that Kim’s view of the constitution of mental states was the most promising one. I found it very appealing that Kim’s view holds that the only truly existing phenomena are physical brain processes and events and that mental states (as well as mental causation) exist only as figures of speech or expressions of our ordinary language.

Although I regard Kim’s view as the most promising theory regarding the constitution of mental states that the philosophy of the mind has to offer, it is far from being a correct or even a good theory. By paving the way for type-identical reductions Kim does manage to resolve the mind-body problem – provided that his argumentation is correct. However, I am certain that there are plenty of scholars who would be more than willing to challenge at least Kim’s claim about the preferability of strong supervenience over a milder version of the concept and the claim that this would facilitate reductionism. Futhermore, besides mere technical issues, Kim’s theory does have a couple of serious defects, which are related to the functioning of the human brain. Kim believes that the structural-restriction thesis is enough to circumvent Putnam’s MR. In fact, it is enough, if one is dealing with multiple realization in the sense that it aims to describe physical realizations of consciousness by different systems, organisms, and species. By concentrating on the physical realization of mentality within the human brain Kim does not have to worry about how consciousness might be implemented by the Martian brain or computerized intelligence. Nevertheless, the problem is that multiple realization is found to occur even within the human brain: the human brain has several, independent physical realization bases for the same mental states. Therefore, Kim’s theory falls subject to criticism based on its denial of disjunctive properties.

Finally, there is the issue of qualia and subjectivity: all conscious states are subjectively experienced and they are found to have a qualitative, ”felt” aspect. The existence of this phenomenon has generally been regarded as evidence against reduction: since mental states have qualia and physical states lack it, mental states cannot be identified with or reduced to physical states. However, this is not necessarily the case. There is no reason to believe that the qualitative content of conscious experience would in some way be beyond the reach of neuroscientific theory. The fact is, though, that a reductionist theory does have to come to terms with this phenomenon. Qualia require explanation, and this is something Kim’s theory cannot provide. The claim that mental states are only figures of speech is not sufficient.

If Kim's theory is really the best that the philosophy of the mind has to offer, the whole area of research has some serious catching-up to do. In the following chapters thorough descriptions of the neurophysiology of hearing and the auditory processing of non-conceptualized sounds will be presented. It will become evident that philosophers of the mind have disregarded the neural sciences far too long. It will be demonstrated that not only Kim's conception of the physical realization of mental states but also most of the current philosophical notions related to the constitution of mental states are in contradiction with the presented neuroscientific facts. Finally, a new kind of theory regarding the constitution of mental states will be presented. The theory will be based on and supported by a vast array of neuroscientific findings.

Chapter II

The Neuroanatomy and Neurophysiology of Hearing in Humans

The purpose of this chapter is to explain how the human brain produces mental states from auditory sensory stimuli. I am going to outline a model of the auditory system, which presents the most important subsystems and subcortical pathways contributing to the neural coding of sensory inflow. I will also consider in detail how travelling sound waves get transduced into electrical signals, and discuss in general how neurons transmit information. The intention of this description, therefore, is to propose the idea that different parts, or subsystems, of the auditory system are structurally and functionally specified, and that their tasks and capacities depend on their physical properties.

Before proceeding a few preliminary remarks ought to be made. Firstly, in the following discussion auditory sensations are strictly confined to sounds that have no semantic content. The auditory system described is meant to give solid grounds for a philosophical thesis. However, this task is compromised if one muddles the discussion with confusing issues concerning the origins of language and meanings. Even though some progress has been made in this area of research during the past few years (see e.g. Aaltonen et al. 1993; Aulanko et al. 1993; Deahaene-Lambertz 1997; Kraus et al. 1997; Näätänen et al. 1997; Pihko et al. 1997) the problems linked with the neural coding of meanings still remain too complex and controversial to be tackled here. So, the focus is on what is known rather than on unanswered questions about language and meaning.

Secondly, it is important to remember that the present knowledge of the anatomy and physiology of the auditory system relies heavily on animal studies. Although the general principles are thought to be similar in all mammals, including humans, there is always the possibility that our notions might be incorrect and need to be refined in the future. Let us take the functional anatomy of the auditory cortex as an example. This has been studied mainly in cats (see Webster & Aitkin 1975; Imig & Adrián 1977), because of easy accessibility of their auditory cortex which is displayed on the surface of the brain. There is, however, no certainty as to what extent the observed neuronal responses of, for example, anaesthetised cats will correlate with those of awake human beings. Though, it should be noted that there are advanced studies on the functional anatomy of the auditory cortex in awake monkeys (see Javitt et al. 1996). In monkeys, the auditory system is structurally and functionally very similar to that of human beings. Therefore, these new studies have provided more accurate knowedge and decreased the possibility of misinterpretations.

Another closely related issue is also worth mentioning and that is the question concerning the principles of neuronal organization. Much of the cellular level research has been devoted to studies on invertebrate animals (Hoyle 1975). The main reason for focusing on invertebrates, or organisms of ”lower” forms, is the structural and functional complexity of the vertebrate nervous system, which has so far resisted comprehensive analysis. Even though some detailed structural-functional analyses made at the cortical level are available (see e.g., Szentágothai 1978), and a better understanding of the functioning and the properties of a single neuron has been reached, the underlying problem still remains: How do larger columns of neurons (such as the human auditory cortex)

function, and can their properties and capacities be reduced to the physical structure and properties of a single neuron? Detailed answers are yet to be found.

This is certainly a problem that should not be sidestepped, for it reveals a crucial defect in modern brain research methods. Present knowledge about the neural basis of human information processing relies heavily on *cognitive psychophysiological* research. This new branch of neuroscience is dedicated to uncovering the neural correlates of cognitive functions by using physical measurement techniques (Donchin et al. 1978). The most widely used method has so far been the recording of *event-related potentials* (ERPs), which are electrical responses elicited by the brain whenever it is processing either external stimuli or some internal event. A single ERP wave can be discerned from a constantly ongoing electroencephalogram (EEG) by means of averaging technique. Usually several specifically localized generator processes contribute as subcomponents to the resulting ERP. These may be totally or partly overlapping in terms of time.

However, it is a well-known fact that only some of the largest ERP components can, at this point, be distinguished from one another. For instance, Näätänen and Picton (1978) have hypothesized that at least six different component processes contribute to the N1 wave (the most prominent deflection peaking about 100 ms after stimulus onset). In addition, they admit that their suggestion is inconclusive, and that further research is needed to clarify the issue. This is a telling example of the fact that only the most macroscopic of the processes generating ERPs are known, while the organization of functional systems underlying larger components on microscopic levels remains totally unknown.

At least Pritchard (1981, p.501) and Näätänen (1992, p.83) have acknowledged this explanatory gap between experimentally observed phenomena (i.e. brain waves) and their physical correlates (i.e. specifically localized generator processes). Näätänen especially stresses that, as research methods improve and provide a sharper resolution, the ERP components currently known will probably become divided into subcomponents and the generator processes will be reconstructed in more detail.

This sort of implicit belief in scientific progress, however, does not dispense with the problem. Researchers holding the view which assumes the capacities of different subsystems are dependent on their physical properties, ought to be particularily troubled by this. The difficulty arises from the fact that all levels of organization are of importance in explaining cognitive capacities. If all the generator processes, or subsystems, contributing for instance to the neural coding of sensory information cannot be uncovered, a complete description of the auditory system is not possible. On the other hand, even if one could discriminate each of these subsystems (which is presently impossible), this still would not guarantee a full understanding of brain functions. It has been conclusively shown that columns of neurons have capacities and properties non-reducible to the capacities and properties of a single neuron.

Despite all the problems and obscurities just mentioned, I am going to propose the idea that all the capacities and properties of every single subsystem of the auditory system are determined by their physical properties. I am convinced that, although present knowledge cannot give a precise account of how groups of neurons work and interact with each other, a strong case for the physical property-dependence of neural functions can still be made. After all there is no evidence indicating that capacities of neuronal populations would depend on other factors than physical properties and structure. There is evidence that voluntary switches of attention may

have drastic affects on the functioning of the auditory cortex (Alho et al. 1999). However, it will be shown that the conscious processes which bring about these attention-switches are nothing but neuronal occurences: as such their capacities are dependent on the same kinds of physical properties and structure as the subsystems which they affect.

1. The Neural Coding of an External Sound Begins in the Inner Ear

Sound is produced by vibrations such as the movement of guitar strings or vocal chords. The vibration of a sound source initiates pressure changes in the surrounding air, which are sent out from the source as a pressure wave with alternating peaks and valleys. The properties of a pressure wave form the basis of sound perception: the frequency of the wave determines the pitch of the sound and the amplitude the loudness of the sound. The simplest sounds (e.g. the sinusoidal tones produced by an oscillator) consist of only one frequency with one static amplitude. However, most of the ordinary sounds perceived by the human auditory system – such as speech, music, and environmental noise – are significantly more complex. Their frequency, amplitude and phase change from instant to instant. In addition, these ordinary sounds usually consist of several simultaneous frequencies. (see e.g. Summerfield & Culling 1992; Rosen 1992).

When the pressure wave reaches the ear, it travels through the external ear canal *(external auditory meatus*) arriving at the ear drum (*tympanic membrane*) which intervenes between the middle ear chamber and the external auditory canal. The pressure wave causes the eardrum to vibrate, and the vibration is then in turn conveyed into a small tympanic cavity, the middle ear. The middle ear contains a series of three small ossicles (*malleus*, *incus*, and *stapes*) of which the malleus is attached to the tympanic membrane. The two other bones transmit the vibration of the malleus to an opening in the cochlea, the oval window, which induces pressure changes in the fluid-filled (*paralymph*) cochlea of the inner ear. (Kessel & Kardon 1979, p.106)

Both the external and the middle ear have an essential role in the production of sensations from auditory stimuli. For instance, without the intermediation of the middle ear the pressure wave would hit the fluid at the oval window directly. The fluid has a much higher acoustic impedance than air, which means that most of the sound energy would in this case be reflected. In consequence, the minimal sound pressure required for the production of auditory sensations would have to be increased (Kelly 1991, p.482). The external and the middle ear have therefore much to do with the efficiency and the range of hearing.

2. The Brain in a Vat Example and the Limits of the Auditory System

Neural encoding of the auditory stimulus does not start until the pressure wave reaches the inner ear. This casts serious doubt on whether the external and the middle ear should be included among the components constituting the physical realizer of auditory sensations. As any reader familiar with contemporary writings on consciousness and mind-body relation already knows, the term ”physical realizer” simply refers to a system that can produce or

realize various kinds of mental states by undergoing physical processes peculiar to it. This physical realizer might be some highly developed nervous system (in our case, the human brain, or more precisely, the auditory system), but any system of an inorganic nature (such as certain machines developed by AI research) will equally well qualify. Of course actions of such machines need to fulfill the established criteria for conscious activity. In philosophical literature the ”physical realizer” of mental states has typically been characterized rather vaguely as a certain brain state or event. In the case of auditory sensations, we can apply this notion more accurately by stating that the realizer of mental states is a series of neural processes participating in the neural coding of auditory stimuli. The problem with this description is that it leaves the external and the middle ear out of the realizing base for auditory sensations. I mentioned earlier the crucial role of these ear components in sound production. Now we need to find out whether or not it is possible to have auditory sensations, if we rule out all non-neuronal processes from the auditory system. In other words, the question that needs answering is: should the external and middle ear be considered as *necessary* parts of the auditory system?

This question brings to mind two issues often discussed in the philosophy of the mind: the phantom-limb phenomenon and the imaginative brain in a vat example. Usually these issues are dealt with simply as matters of curiosity, but in our case they offer us useful assistance in setting the parameters regarding the physical realization of auditory sensations.

Phantom-limb experiences have been reported in patients who have gone through amputation. After the operation some of the patients can still feel pain in the missing limb. The pain is thought to be caused by the chronic overactivity of dorsal horn neurons (see Carlen et al. 1978). The importance of these phenomena lies in the fact that they can be interpreted as corroborative evidence of J.J.C. Smart’s (1959) influential claim, stating that there are no such things as ”slight” or ”stabbing” pain *per se*. According to Smart there are only neural processes that produce mental states such as ”slight pain” or ”stabbing pain”. In summary, a person need not have a leg in order to feel pain in it, for it is the brain that ultimately creates pain sensations as well as all the other sensations.

The idea that neural processes could function as the sole producers of sensations has been extended to its limits by the brain in the vat example. This example is a thought-experiment, which has its roots in one of Descartes’ classic sceptical paradoxes, namely in the ”evil demon” argument (see Descartes 1984, p.14). This modernized version plays with the possibility of an imaginative situation, in which the brain is removed from the skull and preserved in liquid. All the nutritional and other requirements of the brain are fulfilled in order to maintain its vital functions. In other words, this arrangement keeps the brain alive. The most intriguing part of this example, however, is not only in the fact that the life processes of the brain artificially maintained. The sensory inflow of sensations is also simulated by manipulation of modality-specific brain areas. The consciousness, or person, living ”inside” the brain never notices the difference. The person still feels that he is attached to a body. He can go out for a walk in the park, hear the birds singing, see the setting sun, run, trip and hurt his knee. The person can live in this illusory state for the rest of his life never finding out, that in reality, all that is left is the organ of the brain full of wires and tubes, sunk in a pool of preserving liquid (for different variations, see Unger 1976; Putnam 1981, pp. 5–6; Nozick 1982, p.167; Pollock 1987, pp. 1–3).

The brain in a vat example is only a logically possible arrangement. Although it is not feasible in practice (at least not at the present time), the proposed idea of neural processes as the source of sensations and an empirically observed world is not simply a product of imagination. For instance, it is a well-known fact that some sounds can be initiated from neuronal origins. These phenomena, termed *otoacoustic emissions*, were first explicitly described by Kemp (1978). Otoacoustic emissions are based on a reversed transduction process, where the movement of hair cells causes a fluid wave that displaces the foot plate of the stapes. The middle ear ossicles convey the vibration to the tympanic membrane setting it in motion. The vibration of the tympanic membrane can then be recorded in the external ear canal. The subject himself cannot normally hear otoacoustically emitted sounds. There could, however, be a connection between these emissions and *tinnitus*, a constant ringing in the ear (which is usually caused by irritation of the auditory nerve). Even if this were the case, otoacoustically emitted sounds could be interpreted at best as noises, not as meaningful sounds.

Instead of mere noise, Penfield and Perot (1963) discovered that meaningful auditory sensations such as voices or music were elicited in epilepsy patients, when their temporal lobes were electrically stimulated. Experiments were carried out during surgery in the hope of improving the patients' condition. However Penfield and Perot's results are unconvincing. The elicited responses reported by patients were mostly hallucinations, or dream-like states, where subjects heard, for instance, familiar and even recognizable songs or spoken sentences. This indicates that electrical stimulation of auditory cortex could not be used for the artificial production of complex external auditory sensations. The method used probably elicits illusory states by using the "raw material" of sensory memory, which makes it easy for the subjects to discern illusory sounds from real external auditory sensations.

The purpose of considering the phantom limb phenomenon and the brain in a vat example was originally to explore the limits of the physical realizer of human auditory sensations. The presented data suggest that both the external and middle ear should be included into the auditory system. In addition, they should be understood as necessary parts, or subcomponents, of that system. In the case of otoacoustical sounds, it is clear that the role of these subcomponents cannot be ignored even if sounds are initiated by neural activity. The only exception to the rule is tinnitus, which is caused by irritation of the auditory nerve. In theory it is possible that complex auditory sensations (such as environmental noises, and spoken sentences) needed to create the illusory state experienced by the brain in a vat could be produced simply by manipulating nerve fibres. After all, nerves transmit all the relevant information as impulses to end organs, which in turn translate them into sensations (for a review, see e.g. Adrian 1964). However, my understanding of the matter is that this kind of speculation should, for now, be regarded as science fiction.

I think we are justified in concluding, that since external (or any other kinds of meaningful) auditory sensations cannot be artificially produced in humans without the contribution of the external and the middle ear, these components should be included in the physical realization base for human auditory sensations. This means that the notion generally accepted in the philosophy of mind, that brain states and processes are the sole realizers of mental states, is clearly false – at least in the case of auditory sensations. Empirical data show irrefutably that complex and meaningful sounds cannot be elicited by neural origin only. On the other hand, neither can mere neural processes produce mental states from external auditory stimuli. The concept of the physical realization base

is, thus, too narrow and needs to be extended. The content of this extension, in the case of human auditory sensations, should include the external and the middle ear as physical realizers of auditory-evoked mental states. In practice, this makes both components necessary parts of the auditory system.

3. Sound Waves Are Transduced into Electrical Signals by Hair Cells

The brain processes all received information, including auditory sensations, by using stereotyped electrical signals. These signals are ”stereotyped” in the sense that they are virtually identical in all nerve cells of the body. In fact, nerve impulses are found to be significantly similar even in nerve fibres of much less evolved animals, such as whales or even worms. Electrical signals are thus the universal language of nerve cells and the means of their reciprocal communication in all known nervous systems (see Nicholls et al. 1992). In case of auditory sensations, the signals obviously cannot resemble in any way the external world or sounds they represent; impulses are simply symbols, or parts of a code-language, that convey information to different parts of the brain. This means that, at some point, external auditory stimuli have to be transduced into electrical form. The process in question takes place in the cochlea of the inner ear.

The cochlea is a bony, coiled canal consisting of three compartments. Two larger cations, *scala vestibuli* and *scala tympani*, are filled by cerebrospinal fluid, paralymph. Between these two lies a triangular duct, the endolymph-filled *scala media* (also known as the cochlear partition). The scala media is separated from the overlying scala vestibuli by an acoustically transparent Reissner’s membrane. This thin vestibular membrane takes no part in the cochlea’s mechanical functions; its only task is to separate endolymph from perilymph. However, the *basilar membrane* that separates the scala media from the underlying scala tympani has a far more important role in the transduction process. This floor of the scala media contains *the organ of Corti*, which consists of two types of hair cells – outer and inner hair cells (respectively OHCs and IHCs) – and their surrounding supporting cells. On top of the organ of Corti lies the *tectorial membrane*.

The transduction process is based on the physical arrangement between the basilar membrane, hair cells, and the tectorial membrane. Each hair cell extends a bundle of stereocilia into the potassium-filled scala media, the longest of which are attached to the tectorial membrane. When the air-pressure wave finally reaches the cochlea through the oval window, it induces movement in the basilar membrane, propagating a travelling wave which heads towards the distal end of the membrane. The movement of the basilar membrane causes displacements of the stereocilia bundles, which, in turn, initiate depolarization of the hair cells. Voltage changes are produced by the opening and closing of cationic channels. The opening of a channel lets the high-potassium endolymph stream into the hair cell making it depolarize. It has been proposed that channel activation is produced by cilia, which are connected to the top of the cell membrane, and which open the channels when they are mechanically irritated; in other words they are thought to function as ”gating springs” (Hudspeth 1989). The stimulation of the auditory nerve is then accomplished by the modulation of the hair cells’ membrane potential, which, in turn, initiates a transmitter release to the sensory nerve terminals (I will return to this issue later).

Auditory nerve fibres show some spontaneous activity even in the absence of external stimuli. This phenomenon could be explained as the production of background noise, but it has also been suggested that it originates mostly from the hair cell activity of the cochlea (Harrison 1978, p. 413). However, in normal hearing conditions, acoustic nerve responses are initiated by stimuli received from the external world – even if they consist of just noise and tone bursts. Each fibre has its characteristic frequency (CF), which determines the lowest sound frequency, in other words the threshold, for an elicited response. Maximum firing rates of the nerve fibres are in turn intensity-specific: after the sound exceeds the optimal, characteristic intensity level the firing rate saturates (Pickles 1988). Through the firing of its fibres, auditory nerves conduct information flow as signals in the brain stem.

4. Subcortical Pathways and the Multiple Realization Thesis

Opponents of the mind-brain identification have often referred to *the principle of multiple realization* as supportive evidence for their case (see e.g. Macdonald 1989). According to this principle, mental states (and human-like mental life in general) can also be produced by other neurobiological systems similar to the human brain or even by systems of an inorganic nature (computers, AI machines etc.). However, this is just the hypothetical part of the principle; no one has yet developed any such system or created complex synthetic mental life. Rather than logical possibility, the well-established fact that mental states can be realized by different neural processes in different persons offers a much stronger argument for rejecting the mind-brain identification. For instance, observations of patients suffering from various kinds of lesions make a very appealing example. In these patients other brain areas compensated the loss of a functional-specific brain region and took over the cognitive functions of the damaged area (see Glanzer & Clark 1979; LeDoux 1979; Hécaen 1979). So, the crucial question is, how can mental states and brain states be identified with one another, if mental states can be produced by different neural processes even within a single brain? This is the contradiction that is usually thought to undermine the identity thesis.

I think it is worth investigating whether it is possible to overcome this ambiguity. What we need to do is to find out if there is any way to reduce the number of physical realizers to just one. In our case, the basic strategy is to form a model of the auditory system that consists only of necessary parts. This means that every subsystem should be essential to the production of auditory sensations, and that without even one of them the realization would fail.

The obvious starting point for the modeling of the auditory system is hemispheric lateralization. Both hemispheres have independent cognitive capacities that enable them to produce simple mental states individually. In fact, this matter is easily verifiable from so called ”split brain” experiments. One can feed an auditory stimuli to only one hemisphere of a subject’s brain at a time by using the dichotic listening method first introduced by Kimura (1964), or, in the case of visual sensations, the same experiment can be repeated by means of Z-lenses developed by Zeidel (1978; see also P.S. Churchland 1986). The subject will then report either auditory or sensory sensations (depending on which sense is under scrutiny) that are actually processed by a single

hemisphere. It should be noted, however, that the accuracy of these tests can be questioned. The only true ”split brain” cases are lobotomy patients whose cerebellum has been operated on. For instance, in normal test-subjects auditory stimuli always activate both hemispheres. Therefore, too much weight should not be placed on these experiments as sources of information about hemispheric lateralization.

Studies have indicated that the hemispheres' capacities to execute certain cognitive functions vary significantly. For instance, the right hemisphere has a subnormal and much less-developed capacity for discriminating phonetic features (Nottebohm 1979). These kinds of asymmetries are consequences of differences in the structural-functional specificity of the hemispheres. What is meant by this is that the hemispheres are not identical, for they are designed to work together and in a sense to compensate each other as parts of the whole auditory system. However, it has been reported that hemispherical asymmetries in information processing emerge only at the higher level of analysis (see Moscovitch 1979); the right hemisphere can process lower level spectral (i.e. physical) features of auditory stimuli just as well as the left hemisphere. As I emphasized earlier, our discussion is confined strictly to sounds that have no semantic content. Our attempt to model a single-based physical realizer is thus not jeopardized: both hemispheres can be regarded as single-based physical realizers of simple auditory sensations.

At this point we have divided the brain into two separate physical realizers. This is the first step towards a model of the auditory system that contains only necessary subsystems. For the reasons just mentioned, we obviously cannot consider both hemispheres together as a single realizing base: even if large areas of, let us say, the left hemisphere were damaged making it disfunctional, the right hemisphere would still be able to produce auditory sensations (in terms that they would require only lower level information processing). The idea that a single hemisphere could function as the basis for a model of the auditory system seems very promising. If even one subsystem of the model (e.g. the external or the middle ear) fails, the production and processing of auditory stimuli become impossible. This kind of vulnerability offers strong support to the notion that single hemispheres are the right place to start the formation of the desired model.

Impulses can travel through the auditory nerve to the brain stem via a number of different routes that are largely separate. This is illustrated by Figure 1. The fact that the auditory system has several independent ascending pathways seems to confirm the irrefutability of the principle of multiple realizability. However, if we restrict the realization base to only one hemisphere, the necessity condition set for each subsystem is fulfilled.

For instance, signals are conducted by the auditory nerve to the cochlear nucleus (CN) on the ipsilateral side, which in turn projects them to the superior olivary complex (SOC) on both sides. The SOC codes interaural differences in timing of low-frequency tones and disparities in intensity of high-frequency tones, both of which are central functions in sound localizing. The SOC projects the spatially-coded input to the inferior colliculus (IC), which analyses the information sent from the SOC and dorsal cochlear nucleus, passing it on to the medial geniculate nucleus (MGN) in the thalamus. The MGN finally projects the information to the auditory cortex (see Levänen 1996; Evans 1992). If we limit the model to one hemisphere, the CN can project only to the SOC on the same side. As I mentioned earlier, the SOC's function is to code interaural differences in timing and intensity. Now, if the SOC is damaged, the (single hemispheric) auditory system loses its ability to locate sound

sources in space. This indicates that the SOC (like all the other subsystems) is a necessary part of the physical realizer, and that we have hence successfully formed the desired model.

Unfortunately, our single-hemisphere model has one defect which might undermine the whole proposition. The bilateral connections of the subcortical pathways are the key to binaural interaction, which is in turn thought to be the basis of sound localization. Sound localization is performed by the auditory system basically by comparing the auditory stimuli arriving at the two ears (see e.g. Blauert 1983; Deutsch 1988). If we take away these binaural connections, we simultaneously prevent sound localization. This function is very important for humans and all vertebrates, although I am ready to dispense with it, if there are no other means of defending the mind-brain identification. Luckily, this is a step we need not be taken; there is another way to get around the principle of multiple realizability. The solution is the structural and functional specificity of the subsystems. This concept has arisen a number of times during this study, but its content has, so far, been left unexplained. I will now describe the general features of this approach.

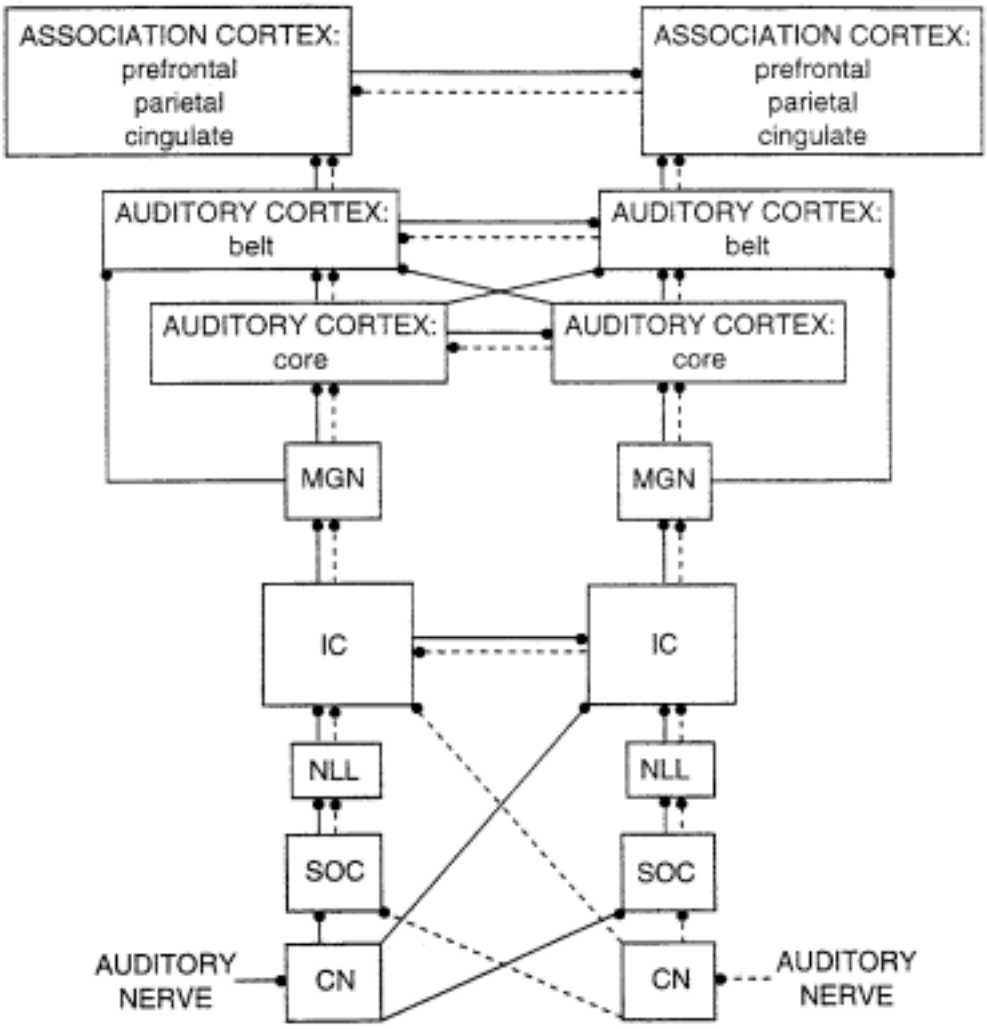


Figure 1. A model presenting the most important ascending pathways from the ear to auditory cortices. CN = cochlear nucleus, IC = inferior colliculus, NLL = nucleus lemniscus lateralis, MGN = medial geniculate nucleus, SOC = superior olivary complex. Reproduced from Levänen (1996).

Later I will discuss a number of specific examples that conclusively show the compatibility of this approach to the auditory system.

By ”structural-functional specificity” we simply mean that a) each subsystem has a certain task to perform, and that b) their capacity to execute a function peculiar to them is determined by their physical structure. The basic idea is that each subsystem has a certain physical (i.e. molecular) structure, which gives it different kinds of capacities. For instance, the tympanic membrane has a molecular structure that makes it flexible, but also breakable if excessive force is directed towards it, and it also has the ability to vibrate. However, by itself, the tympanic membrane has little significance in this world. Nonetheless as a part of the auditory system its capacity to vibrate makes all the difference. As one of the many subparts of the auditory system, the tympanic membrane has the function to vibrate when an air-pressure wave reaches it, and to convey the air-pressure changes to the ossicles of the middle ear. All this is made possible by its structure: if its structure made it any harder or significantly more fragile, its capacity to perform the function in question would decrease or even be prevented. So, the overall picture is such that each subsystem has a specific function to perform, which is made possible by their structure. No subsystem has any function, if they are disconnected from the auditory system (they are basically pieces of material), but their place in the whole gives them meaning as information processors.

But how is this going to help us with our multiple realization problem? Well, the beauty of the concept of structural-functional specificity is its ability to combine the principle of multiple realizability and the mind-brain identification thesis. Firstly, if we consider the auditory system as a system that consists only of subparts whose capacities are completely dependent on their physical structure, we have in our hands a situation where auditory evoked mental states are entirely physical processes. In other words, structural-functional specificity principle contains implicitly the presumption of mind-brain identity. Secondly, the structural-functional specificity does not presume that every subpart of the auditory system is contributing to the processing of auditory stimuli on every possible occasion. For instance, if both hemispheres are included into the auditory system, this does not mean that the cochlear nucleus has to project every time to the superior olivary complex of both sides. If on some occasions only one SOC contributes to the realization of a certain mental state, the other may be considered as *a latent function* of the whole physical realizer. This catches the intuitive idea of multiple realization. A system can use only some of its capacity at a time, which means that only some of the subsystems are active simultaneously, while others are latent: the same system can have several different realization bases inside itself.

The data just presented conclusively show that the principle of multiple realization poses no threat to mind-brain identification. On the contrary, it only strengthens the idea that mental states are produced by systems, that have a large number of simple, structural-functionally specified subparts. Multiple realization is actually made possible by these subsystems, whose unique capacities are on some occasions needed and on some occasions not. Without this active-latent distinction it would be difficult to explain how the brain can gain the capacity to perform versatile functions – even when it is not actually processing anything. The same argument can be repeated, when one asks how the brain is able to recover some of its cognitive functions even after large parts of it have been severely damaged.

5. The Anatomy and Connections of the Auditory Cortex

The human cortical auditory region is located primarily in the superior temporal plane and the superior temporal gyrus. The primary auditory cortex (the core system), located in the transverse gyrus of Heschl, takes care of the elementary analysis of frequency and amplitude. The surrounding and less granulated associative areas (the belt system), that can be further subdivided into a medial and lateral zone, are thought to be involved in more integrative functions, such as sound localization. (Galaburda & Sanides 1980; Forbes & Moskowitz 1974)

The auditory cortex is linked with various parts of the brain by both interhemispheric and intrahemispheric connections. Besides numerous thalamocortical pathways, there exists a descending pathway between the cortex and cochlea that establishes cortical control over the OHCs and the basilar membrane itself (Diamond et al. 1969; Weedman et al. 1994). In addition the pathways of the corpus callosum provide cortico-cortical connections between the primary auditory cortex of opposite hemispheres.

The auditory cortex is functionally organized: neurons are spatially distributed according to their maximum response range to specific stimulus features. For instance, neurons of the cortices are tonotopically organized, which means that each neuron has a certain CF at which the threshold intensity is lowest. Neurons in the primary auditory cortex especially are frequency selective, but also some association areas have cells that have a specific frequency range (Romani et al. 1982; Tiitinen et al 1993). There is also evidence of an amplitopic organization in the auditory cortex. These findings suggest that cells also have a characteristic intensity (CI) to which they respond best (see Pantev et al. 1989). This kind of topical mechanism for intensity coding can be observed by intensity changes that might ”shut down” the active neuronal population and bring an earlier less or even totally unresponsive area of neurons to maximum response.

Further evidence of the structural-functional specificity of cortical neurons is provided by the contralateral dominance of the sensory cortices; the primary auditory cortex responds more efficiently to monaural stimuli at the contralateral ear. For instance, it has been shown that different neurons and neuronal populations in a cat’s primary auditory cortex favour stimuli from only one of the ears (Imig & Adrian 1977). In addition some neurons respond solely to stimulus offset instead of onset, while others might react to both of them (Abeles & Goldstein 1972).

6. The Auditory System is Structurally and Functionally Specified

I have repeatedly emphasized that my description of the auditory system is based on the notion that different brain parts or subsystems are structurally and functionally specified. To put it more clearly, each part of the auditory system has a certain molecular structure, which at the same time both enables and determines its capacity to perform one or several functions needed in auditory information processing. The notion of structural and functional specificity contains implicitly the very strong presupposition that the realizing base of complex and ”felt” auditory sensations can actually be reduced or divided into totally unintelligent, basic physical processes,

such as the movement of ions. Every fact about the neurophysiology and neuroanatomy of hearing I have presented is in perfect accordance with this presumption.

Sounds have to go through numerous alterations and transformations from air-pressure changes into electrical impulses before they are detected as sensations. Each subsystem participating in this complex transformation seems to be perfectly designed to meet the requirements of its task-performance. For instance, the ossicles of the middle ear have just the right composition to react to the vibration of the tympanic membrane. The fluids of the inner ear have exactly the right concentration to enable the cochlea's hydromechanical process for transducing mechanical stimuli into electrical potentials: the endolymphatic fluid is positively polarized (the predominant cation is potassium), which makes K^+ ions stream into electrically less charged potassium-filled IHCs, when their cationic channels open. The structure of auditory nerve fibres has tuned them to be more responsive to characteristic frequencies. In addition, the superior olivary complex has three functionally specified nuclei, and the neurons of the auditory cortex are both tonotopically and amplitopically organized responding to specific frequencies and intensities.

These few arbitrarily selected examples make it clear that structural and functional specifity is the common nominator of the different subsystems throughout the whole auditory system. I believe this argument stands on its own and more detailed description is unnecessary; one has to challenge the present knowledge of the neurophysiology of hearing in order to refute this principle. However, even though empirical facts support the notion that the auditory system consists of several unintelligent subsystems, and that the capacities of each subsystem is determined by its physical structure, there are still few points that need clarification.

First of all, there is a huge gap between the detection of sounds and the movements of electrical impulses or simple physical particles. Therefore it is justified to ask, whether these elementary processes, that in addition could at best be described as passive occurrences in the presence of an initiating cause can really cover the physical realization of complex sounds? I must emphasize that the issue under question here is the passive nature of the subsystems' task-performance. By "passive nature" I mean to describe the ambiguity of the notion that, on the one hand, supposes subsystems to be lower-level information processors, but on the other hand makes their outputs seem like mechanical reactions to input stimuli. References to subsystems' physical property-dependent capacities only reduces the credibility of this notion. Glass has a crystalline structure that ensures that a window breaks, when a stone is thrown at it with sufficient force. However, a view that identifies information processors with shattering glass is not a believable one. I am going to deal with this problem by rejecting the image of subsystems as passive reactors.

Another issue that has not yet been considered in depth is the subsystems' physical properties. I have established that subsystems' structurally and functionally specified capacities are dependent on their physical structure, but yet it is unclear what determines their physical properties. I will try to shed some light on this issue by considering some of the cellular properties of neurons and their origins.

7. Are Subsystems Only Passive Reactors?

Can passive vibrating cover the analysis and realization of sounds? The answer is clearly no. Even the human hearing range contains such high frequency-levels that simple electrical or mechanical resonance (used in lower vertebrates) can no longer meet the requirements of sound production. Instead of passively reacting, the human auditory system has to actively analyse the spectral features of sounds through interaction between receptors and sensory accessory structures.

The inner ear's role in information processing makes it a useful example of the activity of the auditory system. For almost 30 years conceptions of the frequency analysis of the inner ear were dominated by a paradigm originated by von Békésy (1960). According to his description of a travelling wave, the basilar membrane undergoes only passive linear vibration. Later research has conclusively shown this notion to be false; the BM is not only highly non-linear, but also very sharply tuned (Johnstone et al 1986). Nevertheless, von Békésy was partly right in one sense: the BM does have a passive mechanism. When the cochlea's fluids induce movement of the basilar membrane, it results in a wave travelling towards the distal end of the membrane. The basilar membrane's structure is functionally specified in such a way that every location in the BM has a characteristic frequency (CF). Travelling waves initiated by high-frequency sounds reach their maximum amplitude near the base of the membrane, while low-frequency tones reach their amplitopic peak closer to the apex.

But in addition to passive mechanics, the cochlea also participates actively in the frequency analysis. The key element of an active cochlea is different innervation patterns between OHCs and IHCs. Approximately 90–95% of the afferent neurons terminate in IHCs, and the same fibres do not innervate both IHCs and OHCs (for an overview of efferent innervation, see Warr & Guinan 1979). The role of the IHCs is to function as the sensory receptor in the hearing organism. OHCs can function also as sensors, but it has been suggested that their main function is to provide feedback elements (Dallos 1992; Ruggero et al. 1992). The basic idea is that the cochlea has a mechanical feedback process, which can either amplify the passive travelling wave or produce "negative damping", a resistive force that enables the basilar membrane to handle much larger response amplitudes. In both cases the cochlea's tuning is sharpened and the response range increased. OHCs are thought to be behind this active feedback which can feed back energy to the basilar membrane–tectorial membrane system. However, this activity contains nothing but purely physical and mechanical processes. The energy released by OHCs is produced by changes in their receptor potentials (depolarization–hyperpolarization), which in turn is propagated by (mechanical) acoustic stimuli.

The active cochlea provides thus strong proof against the passivity claim that endeavours to exlude all activity outside structural-functionally specified subsystems. The cochlear tuning is clear evidence, that even though subsystems' capacities are dependent and determined by their physical properties, the auditory system can hardly be referred to as a series of breakable windows. The human auditory system is hence living proof of the fact that even systems that consist only of basic physical and mechanical processes can be highly evolved, complex, and able to accomplish tasks that require intelligence.

8. The Origin of Cellular Properties

Electrical signals are the universal language of neurons and their means of communication. Signals are mediated in the brain by action potentials, which are transmitted along axons. Action potentials trigger the release of neurotransmitters in chemical synapses, which, in turn, open selective ion channels in the membrane of the postsynaptic cell. The opening of the ion channels causes current to flow along the interior of the cell thereby generating postsynaptic potentials (PSPs), which are either inhibitory (IPSPs) or excitatory (EPSPs). Although every single detail of cell's physiology is not yet exhaustively understood, the basic principles of neural functions are fairly well established.

As a matter of fact, the versatility and complexity of brain processes do not arise from the functioning of a single neuron. The enormous capacity of the brain is instead made possible by different connections between neurons. There are approximately 100 billion neurons in the human brain, each of which can have at least 10 connections with other neurons, some inhibitory some excitatory. The connections are, in addition, of different strengths ranging on an approximate scale from 1 to 10, which means that there are at least 10 raised to the 100 trillionth power ($=10^{100\,000\,000\,000\,000}$) possible connections between neurons (see P.M. Churchland 1996, p.5). This ought to show that the properties of a single neuron have strong bearings on different brain functions. In fact, the general trend in research has in recent years shifted away from the analysis of the total brain and its larger anatomical subdivisions into studies that concentrate on separate cell types and fractions of nervous tissue. This is one of the reasons I will consider shortly some of the cellular properties of neurons. I decided to focus on ion channels which have the important function of making the propagation of action potentials possible.

Three factors determine whether a particular ion type can pass through an open channel, and at what rate its members are bound to do it. These factors are the ion channel's *gating mechanism*, *permeability*, and *selectivity*. By "gating" we simply mean the process that transmits the channel's state from shut to open. There are three commonly found mechanisms. Firstly, *ligand-gated channels* may open when a neurotransmitter or internal messenger binds to the channel molecule's binding site. In *voltage-gated channels* the triggering event is a change in the electric field across the membrane. The third possibility is *mechanosensitive channels*, whose opening is caused by mechanical irritation or stress applied to the cell membrane. These types of channels can be found in the IHCs of the inner ear. There are also some less familiar gating mechanisms; nevertheless the basic principles are the same in every variation: gating is a process, where change in one part of the molecule produces an effect in a different part of it by opening a pathway to ionic current (for a specific description, see Aidley & Stanfield 1996, pp. 161–223).

As we earlier learned from the functioning of IHCs, the movement of ions is due to their electrical properties. Extracellular space has a higher concentration than the membrane plasma. The ions at the site of higher concentration tend to move away from their own type towards the ions inside the membrane, whose electrical potential is opposite to that of extracellular ions. The rate at which these specific ions pass through an open channel into the membrane (under standard conditions) is governed by the degree of permeability. However, ions cannot move freely through every ion channel. Ion channels can be highly selective: they can be only slightly permeable, or not permeable at all, to ions other than those they are primarily selective for. For example,

Na channels are very permeable to Na^+ ions and less permeable to K^+ , while K channels are very permeable to K^+ ions but not to Na^+. My interest here is to find out why channels allow only some ions to pass through and not others. Which physical properties determine the selection for ions?

There is a lot of ground to cover regarding selective ion channels. Unfortunately, there is no time to dwell on this issue. Hence, I will simply outline a simplified theoretical model, which indicates some of the most important physical properties relevant to ion selection. Let us take voltage-gated, highly selective Na channels as an example. The flux inside the cell is partly determined both by the internal concentration gradient and by the electrical field. This explains why Na channels are poorly permeable to K^+ ions: the electrical properties of K^+ ions give them a much lower degree of permeability than Na^+ ions in an Na channel environment (Hille 1992, p.351). This is, however, not directly related to channel properties but depends on the electrical charges of particles.

However, there are two other factors responsible for the Na channel's selectivity that are entirely determined by channel structures. Besides the electrical potential differences that accompany permeation, the shape and size of the pore set certain criteria for passing ions. Firstly, the pore has to be narrow enough to force the ions in contact with the wall; otherwise they cannot be recognized. Secondly, ionic channels have to have a selective filter. On the other hand, its task is to prevent the permeation of larger particles, and furthermore it is designed to provide oxygen dipoles as surrogate water molecules, whenever an ion has lost some of its contact water (Hille 1992, p.357). The Na channel's selectivity is based on its pore size. For instance, the Na channel's pore will not allow methyl groups to pass through, but it is just wide and tall enough to let aminoguanidine stream into the cell by making hydrogen bonds with the selectivity filter. Now, the final question is, what determines the shape, size, and electrical conductance capacity of ionic channels?

The functions of any cell are determined and also mediated by proteins. Alterations in a cell's functions are a consequence of the synthesis of specific proteins, which are responsible for its specific functions. All the different functions of neurons – the conduction of action potentials, synaptic transmission, and the establishment of specific connections – fall under this description. However, in the case of neurons and the nervous system in general, their functions and capacities are nowadays thought to be mediated by certain brain-specific proteins not found in other organs (Moore 1973). For instance, when the brain was found to be rich in tubulin (a subunit of microtubules not unique to the brain organ), especially in neurite extensions, in axoplasmic flows, in endocrine secretion, and in cell-surface site distribution, it was postulated that tubulin played an important role in nervous functions. The precise function of microtubules is not yet known. However, the two main postulates are that they either function in the formation of cell structure as a supporting element, or they are key components in the motility-transport system of a cell (Shelanski 1973; Peters et al. 1970, p.62). Nevertheless the fact is that whatever the specific function of microtubules turns out to be, it is determined by tubulin's properties, in other words by its ability bind to other molecules and form entities.

All the properties of ionic channels are also determined by proteins. Usually ionic channels are composed of one or more protein molecules, each of which in turn consists of several subunits ranging from four to six tokens. In practice channel properties are dependent on these subunits. Each subunit is actually a chain of amino

acids combined in linear sequence by peptide bonds. Amino acids are arranged in unique sequences giving the protein molecule certain properties, such as structure, an ability to conduct electricity and so on.

9. In Place of A Summary

We are now in a position to sum up our discussion on structural-functional specificity. Our original presumption, reinforced with empirical data, was that the auditory system consists of several unintelligent subsystems, whose specified capacities are determined by their physical properties. We learned that ”physical properties” refer to either molecular structure or cellular properties. Along the way we had to admit that some groups of neurons and subsystems might have capacities which are non-reducible to properties of single neurons, but it was nevertheless made clear what an important role single neurons have in brain functions. It was shown that neurons have characteristic functions and capacities that are mediated by proteins, which can in turn be reduced to amino acid sequences.

The roots of structural-functional specificity go further than amino acids. The whole structure of the auditory system is ultimately reducible to DNA molecules that contain the genetic information. It is DNA that determines amino acid sequences by translating its information through RNA into proteins (for a detailed description, see Aidley & Standfield 1996, pp. 60–68) Though it has been conclusively demonstrated that the environment can alter the nervous system both structurally and functionally (see, e.g. Jouvet 1978), the blue print of the auditory system is stored in these two-stranded polymers. The DNA molecules guarantee that every subsystem has just the right physical composition or cellular properties enabling them to contribute to the physical realization of auditory sensations. DNA is the container and executor of a master plan created by millions of years of evolutionary processes. This plan is actually a set of instructions for building complex organic systems that can process auditory information and produce sensations.

Chapter III
Auditory Information Processing

I have so far considered consciousness from two rather narrow perspectives. First of all, my main interest has been in outlining the mind-body problem and presenting the current debate with its major positions. Basically, these discussions have concentrated on the ontological status of mental states and on the relation between the mental and the physical. I have also introduced some reflections on what effects the special features of consciousness (such as intentionality, qualia, and mental causation) have on these issues. In addition, the viewpoints of action theory and folk psychology and their roles in the mind-body debate have briefly been dwelt on. However, I must repeat that the main interest is not in conscious activities or human awareness in general. Instead, the study concentrates on the constitution of mental states (or events, properties, and predicates) and their relation to the physical domain. Secondly, I have presented neuroanatomical and neurophysiological bases for what seems to be a very limited part of mental life; that is, auditory sensations. Moreover, I have restricted the auditory sensations under consideration to only very simple, meaningless sounds, which have no semantical content. In other words, we are only dealing here with small fractions of human mental life, and the brain functions under scrutiny cover only a minor part of the capacity of the brain.

Since most of what is usually thought to constitute consciousness, or to be related to conscious acts, has been left out, the restrictions evidently seem to diminish the value of this work. Furthermore, I have not yet presented any well-articulated reasons or arguments for why these restrictions were intially made. Finally, the fact that I have not tried to provide any connection between the thorough reflections concerning the ontology of mental states and the purely neuroscientific descriptions of the auditory system raises some doubts about the soudness of my approach.

It is for these reasons that I have assigned a twofold purpose to this chapter. Firstly, I wish to erase doubts and obscurities by clarifying my intention and by explaining the link between the different perpectives I have taken. The first three sections of the chapter are devoted to this task. In order to explicate and justify my position, I will turn to *cognitive science* and *cognitive neuroscience*. I will show that these areas of research contain some paradigms which facilitate and give considerable support to my efforts at formulating a reductionist theory based on empirical neuroscientific evidence.

The second purpose of this chapter is to complete the description of the auditory system started in chapter II. The rest of the sections present a model, which explains how auditory information is processed in the human brain. After this chapter, we will have hopefully reached a comprehensive understanding of how the brain transforms auditory stimuli into auditory sensations. At least a fairly substantial – if not complete – knowledge of the neuroanatomical facts and neurophysiological processes that participate in the production of sounds will be explained. This will give us the chance to compare the presented neuroscientific facts with a few central notions from current theories of the mind. The following chapters will then be assigned to evaluating the accuracy and the validity of these theories.

1. Mental Representations and the Demand for an Inner Representational System

Philosophical talk about the constitution of mental properties and technical reports about electrical potential measurements of some cortical-level neurons are in danger of obscuring the more general notion of what consciousness is or what conscious acts entail. What I am striving for here is not some ontological disappreciation of special features of consciousness. Neither am I complaining about a huge gap existing between psychological and neural theories. Instead, I wish to take a purely practical point of view. Let us think of the simple fact that through perception we are constantly aware of reality outside consciousness. Our sense organs feed the brain with an unending flood of information, which helps us to construct and retain an adequate picture of the surrounding environment. How is this possible?

One influential paradigm originates from Franz Brentano, who considered the same question and came up with the answer that it was the *intentionality* of consciousness that made it possible. Brentano claimed in his classical *Psychology from an Empirical Standpoint* (1874) that it was the ”object-directedness” or ”aboutness” (i.e. the ability to be directed towards material objects of reality outside itself) characteristic of consciousness that facilitated our knowledge of the outside world. He also concluded that since material reality lacked this capacity, mentality had to be immaterial. However, the traditional framing of the question has gradually taken a new form. Nowadays, the question of how the mental can be ”about” things is not considered to be particularly interesting. The more intriguing issue is how the mental can present objects to itself. (see Hills 1981, p.12)

The knowledge we have gained regarding perceptive faculties and the movement of light rays has made it less interesting to contemplate how consciousness can be directed towards, for instance, a tree in the backyard. After all, we know now that seeing does not necessarily require any active contribution from the one who does the seeing. Nevertheless, what is of interest in this situation is the question of how consciousness forms a mental picture of the tree in the backyard; that is, how does the mind represent things as being one way or another?

Perhaps this shift of focus should be explained in a more precise manner, for I am not sure that the usefulness of the new approach, which concentrates on mental representations, has yet become apparent. There is, indeed, an alternative way to present the difference between the problem of the intentionality of consciousness and the issue of mental representation. The intentionality of consciousness can be expressed in another way by stating that many mental properties – especially those that express propositional attitudes (believing, desiring, wanting, and so forth) – are *relational properties*, which appear to relate people either to *propositions* (see Field 1978, p. 79) or *sentences* (Carnap 1947). The new approach, which is less interested in the concept of intentionality, does accept that propositional attitudes should be analyzed as relations. However, it rejects the assertation that the other member of this relation is either a proposition or a sentence. Instead, it claims that propositional attitudes are relations between organisms and their internal representations (see Fodor 1978).

The differences between these approaches are clearly evident. Let us start with the first alternative. The claim that propositional attitudes are relations between people and propositions states only that there is a set of ordered pairs whose first members are people and whose second members are propositions. This can be exemplified by the set-theoretic definition according to which pain is a set of people, who are said to ”be in pain” or to ”feel pain”. Undoubtly, much could be said about the validity of this approach. Because of the lack of space

here, I will settle for repeating the observation made by Hartry H. Field (1978, p.80) who argues that the description makes pain a purely set-theoretic entity and misses the notion that pain is suppose to be a property and, in addition, something mental.

The second alternative, which was proposed by Rudolf Carnap in *Meaning and Necessity* (1947), is no more promising. The claim that propositional attitudes are construed relations between people and the sentences they are disposed to utter can be dismissed by using two rather straightforward arguments provided by Jerry Fodor (1978, pp. 53–54). Firstly, propositional attitudes are not behavioral dispositions, which undermines the claim that beliefs correspond with belief-ascribing sentences or tokens of them. Secondly, it is highly plausible to think that a person (or a thing) who does not know any language at all can have beliefs. Hence propositional attitudes do not necessarily require language skills. Animals are a living proof of this: they do have cognitive faculties and beliefs, even though they are not as evolved as those of humans.

I believe that the evidence against Carnap's view is so overwhelming that it ought to be put aside for good. But what about the proposition idea? Probably the only way this first approach could be salvaged would be to interpret it as a version of the causal/functional role theory proposed by such writers as David Lewis (1966;1970;1972) and D.M. Armstrong (1968). A functionalist interpretation would thus facilitate the description of beliefs independent of both *language* and of *inner representations*. For an organism to have a certain belief is to say, that it has a state which is causally connected to certain inputs, outputs, and other mental states, and which plays a certain causal role in the organism's psychology.

But it is precisely in this form that the true weaknesses of the approach is revealed. The easiest way to illustrate my point is to repeat the initial question: how does the mind represent things as being one way or another? The answer is that there is simply no way of distinguishing the difference between alternative ways of saying the same thing. As Fodor (1978, p.60) has noted, propositions are pure, formless content which, as objects of propositional attitudes, provide no means by which the different aspect of the same attitudes could be distinguished. And we have so far considered only propositional attitudes. What about other mental predicates? For instance, those involved in perceptual activities? I have the feeling that the benefits of the newer idea are becoming evident. At least, it can easily circumvent the problems with propositional attitudes: for an organism to have a proposition is to stand in a certain (causal/functional) relation to an internal representation which expresses the proposition. But how does this approach apply to other mental states? Apparently we need to take a closer look at the issue.

In philosophical debates the way the mental represents things has been seen to closely resemble the way language represents things. The dominant idea in philosophical literature is that there has to be a system within which mental representations take place. According to the most radical interpretations, this system is thought to constitute a genuine language of thought with its own ”vocabulary” and ”grammar”. The notion of inner language is, though, not entirely new. For instance, one of the earliest participants in the modern discussion, Peter Geach, claims in his *Mental Acts* (1957) that some elementary ideas regarding the analogy between thought and language can be found from the works of medieval writers such as William Ockham and St.Thomas.

However, even though present discussions about inner language do share some vague historical roots and the belief that mental representations need to be supported by some innate (and probably language-like) system,

there are at least two major issues in respect to which the views apart. Firstly, what is the constitution and the status of the language of thought? What is the relation between the language of thought and natural languages such as English?

One of the strongest views has been presented by Jerry Fodor in *The Language of Thought* (1975) in which he insists that an adequate psychological theory cannot be formulated without the postulation of internal representational systems. Fodor claims, that the structures and features of these systems are so similar with those of natural languages, that it is justifiable to talk about a language of the mind in which thoughts take place. The radical nature of this view is based on the fact that Fodor presents a complete (though speculative) description of the language of thought, and that he holds this language as able to express the same things that are communicated in any natural language. So, Fodor's conception is much more than a suggestion meant to realize the loose idea of a framework for mental represensations.

Gilbert Harman (1970) has argued against a clear cut distinction between natural languages and a language of thought, which is implied by Fodor's notion. Harman does not believe that learning language is about converting elements of a distinctive, inner system of representations for thoughts either into the complete sentences of a natural language or some instances (e.g. tokens and deep structures) of it. According to his own "incorporation view", knowledge and usage of a natural language requires the ability to think in that natural language. Put plainly, learning a natural language is accomplished by incorporating the natural language into the inner language of thought.

Daniel C. Dennett (1977), in turn, has doubted Fodor' s claim that psychological theory can be explanatory enough only if our intentionalistic calculations (e.g., to take certain actions given certain beliefs and preferences) is seen to have a syntactic structure. Dennett takes this to be at best an interesting empirical claim, which has not been adequately proven.

The views above are only some of the positions taken on the issue of internal representation and as such represent a small fragment of the debate on the language of thought. However, even from this very limited survey it becomes apparent that the notion of inner representation and the idea of an inner system in which these representations take place might not be so promising after all.

2. Outlining the Main Problems with the Concept of Representation

Issues related to consciousness tend to be troublesome, and the way the mental represents things is no exception. The requirement of an inner representational system is solely sufficient to exemplify my point. After all, the existence of a language of thought is still merely a suggestion, and even if one accepted that such a language existed, there is no certainty as to what its exact nature and form might be. However, besides all the obscurities and uncertainties related to inner representation, I believe there to be even more severe problems which ultimately undermine the whole debate. These difficulties are raised by the problem of locating the origin of meanings.

Much of our discussion has dealt with propositional attitudes, which always have some semantic content. For instance, the possession of the simple belief that it is going to rain tonight requires that the person who has this belief understands what such terms as "rain", "tonight", and so forth mean. The situation is no different in

case of sensations: if someone takes even a peek at the tree in the backyard, the person in question has to possess at least an elementary knowledge of the concepts of ”tree”, ”backyard” etc. These semantic contents are precisely the core of the problem. We do not have any uncontested notion of what meanings really are or what is the meaning of ”meaning”. Correspondingly, we can detect some occurences in the visual cortex of a person who sees a tree in the backyard but we do not know too much about how or where the brain produces the concept of ”a tree in the backyard” –and what we do know is very elementary. Furthermore, Hilary Putnam (1988) has proposed a theory of meaning which denies the possibility that meanings could be explained in terms of the neural sciences. Putnam claims that the reference between words and objects is socially determined: meanings are not in the brain but in the world. It is quite evident that these are issues that will not be resolved in the near future.

If one takes a look at the debate from a greater distance, it is possible to discover that the problems related to meanings are actually implied by the very concept of representation. Antti Hautamäki (1997, p.26) has justifiably pointed out that talk of mental representations run the danger of provoking a copy theory of knowledge; that is, a view according to which the relation between the consciousness and reality is seen as a correlation between concepts and reality or between beliefs and states of affairs. Now, if knowledge is apprehended naively as a mental picture of reality, we are not only stuck with meanings but we also have to come to terms with a few additional problems.

The postulation of a correlation between some mental image and an object of reality is against all known facts. For instance, mental representations do not seem to resemble in any way the objects they represent. Sounds do not have anything in common with auditory sensations which are suppose to represent them; sounds are basically air-pressure changes which are transducted into electric potentials in the cochlea of the inner ear. In fact, the copy theory of knowledge is epistemologically unacceptable due to the much simpler fact that human observatory faculties have their essential features and limitations. Human perception is based on the detection of electromagnetic waves of different altitudes and not on the sensing of, for example, temperature changes. In the same way the human auditory system has the capacity to process only sounds within a certain frequency range. Finally, the physical size of a human body places us in the middle of the macro and micro worlds, which decisively limits our knowledge of the world. We may only think of the troubles we have encountered in the areas of quantum physics or cosmogony to realise this. The bottom line is that human knowledge is always subjective –and I do not mean that knowledge is knowledge of something to somebody only from the point of view of an individual but also from the point of view of a species.(see e.g. Waugh 1995)

Is there anything we could do to get around these difficulties or do we have to submit to another philosophical debate that might go on for a whole century? In the following, I am going to show a couple of manoeuvres by which the problems related to mental representations can be tackled. In doing so I will make use of some key ideas from cognitive neuroscience.

3. Information Processing in terms of Cognitive Neuroscience

We have so far considered the way the mental represents things from a quite a narrow angle. In order to understand and resolve the problem of an inner representational system and other difficulties related to mental representation the issue has to be put into a larger perspective. A brief look at the fairly new discipline of *cognitive science* will help us to see more clearly the more profound questions that hound the issue of mental representation.

The creation of cognitive science was initiated as a result of growing dissatisfaction with the modern philosophy of the mind which was thought to leave out consciousness itself. What the critics had especially in mind were the *qualitative* aspects of conscious experience, *subjectivity*, and *intentionality*. Functional characterizations (which had already established a prominent position in the mind-body debate) seemed to fail to capture these essential features of consciousness. For instance, *computer functionalism* or *strong artificial intelligence* of the 1950s saw the ontological relation between the mind and the brain as identical with the relation between a computer and its program. The problem with these sorts of abstract characterizations of consciousness was that two systems might have appeared to obtain the same functional state but only one of the systems actually had the qualitative, ”felt” experience related to, for instance, smelling of a rose or seeing something blue (see e.g. Revonsuo et al. 1994).

The discipline of cognitive science was meant to correct this error by openly fusing elements from philosophy, AI research, psychology, and the neural sciences. It generally holds that cognition (lat. *congnoscere*) is a form of information processing in which sensory inputs are transformed into mental representations. Various computational processes can in turn make use of these representations in the creation of higher conscious functions such as thinking and forming beliefs. This notion was highly influenced the model Jerry Fodor presented in*The Modularity of Mind* (1983)(see also Putnam 1984; Marshall 1984; Shallice 1984; Jackendoff 1987). In spite of its noble goals, cognitive science has so far proved to be simply an interesting branch of consciousness research which has not yet made any real breakthroughs. Nevertheless, it has helped to clarify some of the more practical problems linked to mental representations and and to identify new obstacles to understanding the relation between consciousness and the brain in a more precise way.

One of the reasons why cognitive science has not managed to provide any better starting point for the research of consciousness is that it is far from being a uniform approach. This is only a natural consequence of its central idea which is to cross over and combine different scientific practices and research methods. However, there are some major paradigms or lines of thought to be found which will help us to organize the disparate field of cognitive science. Although the overall aim of cognitive science could be characterized as the study of cognition in terms of physical symbol manipulation (see Garfield 1990), two separate traditions can be abstracted within which the program is carried out: *philosophical* and *naturalistic*.

The ”philosophical” tradition (also known as the ”continental” tradition) has focused its interest mainly on epistemological questions and on issues recurrently reflected in the philosophy of mind. It relies heavily on the twentieth century’s continental psychology and philosophy (e.g. phenomenology, psychoanalysis, and postmodern cultural studies). This tradition is quite incompatible with the views I have so far presented and will

be presenting, for it is antireductionist and nonmechanistic in nature. Furthermore, it stresses the importance of feelings, actions, and the holistic and social aspect of man (Hautamäki 1997, p.27). For these reasons, I will not consider it any further.

However, the ”naturalistic” tradition, which was developed mainly in Great Britain and North America, is a far more interesting option. The reason for my enthusiasm is that the naturalistic tradition takes cognitive science to be above all an empiristic effort to explain cognition (see e.g. Gardner 1987). In practice, this means that empirical methods and formal models have such a central role within this tradition that cognitive science could –in extreme cases– be interpreted as a member of natural sciences. In this latter form cognitive science is also unavoidably reductionist. Naturalistic tradition has four basic directions, which are *artificial intelligence* (see Boden 1990), *cognitive neuroscience* (see Kosslyn 1994), *connectionism* (see Smolensky 1988), and the most recent invention, *the study of consciousness* , which has wrestled with so called ”global workspace theory” (see e.g. Baars 1994).

I promised earlier that our consideration of cognitive science would help to clarify the problems related to mental representations or even resolve some of them. It is fair to say that the rejection of philosophic tradition does in fact provide a possible way to get around these difficulties. The tradition stresses holism and the social aspect of consciousness; that is, that consciousness research should not disregard language and culture. In practice, this means that the research would be burdened with meanings for good. By rejecting this approach and by favouring the naturalistic tradition one can avoid this frustrating situation which evidently results from philosophic tradition.

Nevertheless, I have already implied that cognitive science is not capable of solving the troubles with mental representations. Of course, this conclusion might seem premature, for I have not presented any thorough treatment of even the most influential versions of cognitive science. However, there is an argument which shows that the problems with mental representations, in fact, originate from the very structure of the program of cognitive science. Cognitive scientists generally hold that cognition consists of three different levels of analysis (see Revonsuo 1997, p.160): a) semantic (content), b) syntactic (algorithmic), and c) mechanistic (structural implementation).

The last level of organization is virtually unproblematic. Cognitive science – like all the other prominent philosophic theories of mind – is driven by functional characterizations. In other words, what is important is the abstract functional organization of the system – not necessarily the specific material which realizes it. For instance, aritificial intelligence and connectionism both take computer programming to be a paradigmatic model for consciousness. This only goes to strengthen the point that the focus of scrutiny is laid on abstract, functional definitions of organizations and not on the physical structure itself. Since it is accepted that consciousness can be realized in various systems of different structures and compounds (at least in principle), structural implementation might not be the biggest bone of contention.

The first two levels of explanation are, in turn, much more troublesome than the mechanistic level. First of all, some cognitivists – Jerry Fodor (1975) among them – have claimed that since cognitive processes can be explained by computer models, they ought to be taken as a serial or linear processing of symbolic representations. However, I already made it exhaustively clear in chapter II that information processing in the brain is parallel in

nature. Secondly, the presupposition that cognition could be imitated by computer models and programs suggests that cognitive processes are essentially mechanic; that is, that they are computable and algorithmic like computer programs. Yet again there has been some evidence that mental processes are, instead, non-algorithmic (Penrose 1989; see also Hautamäki 1997, p.30).

These are still only minor details, which can probably be worked out with a few adjustments. The true cause for the downfall of cognitive science is hidden in the semantic level of explanation. This level of analysis makes a direct reference to semantic contents, which naturally designate the meanings of mental representations. The consequence is that even if the naturalistic tradition was followed and the social and cultural aspects of consciousness were disregarded, one would still have to confront all the problems related to language and meanings.

I may try to clarify my position with a few additional notes. The generally accepted belief, which is also shared by cognitive scientists, is that natural phenomena are organized hierarchically in nature. This puts researchers in a position where there are certain levels of natural organization and certain levels of scientific analysis (i.e. different scientific disciplines), which are suppose to correlate with each other (see Bechtel & Abrahamsen 1991). The situation can be illustrated by the following classification of the different levels of explanation (see Hautamäki 1997, p.28):

1. consciousness (*introspection*, *every-day psychology*)
2. mental processing (*cognitivism*, *cognitive psychology*)
3. neural processing (*neuroscience*, *biology*, *connectionism*)
4. physical phenomena (*physicalism*)
5. quantum phenomena (*quantum theory*).

Now, everybody can see that good old reductionism is at work here. The basic idea is that the higher levels are reducible to the lower levels, which are thought to be more basic and hence provide more accurate knowledge or more objective explanations of mentality.

Let us consider these levels from the bottom to the top. In recent years a fairly intensive debate has evolved over whether quantum theory has any part in explaining the mind-body relation (see e.g. Perus 1995; Hiley 1997; Globus 1997). Although the discussion has been rather dramatic and interesting, it is doubtful that quantum mechanics would be able to provide a final explanation for consciousness. Anyway, the obscurities alone that are related to quantum theory will guarantee that this is not going to happen in the near future. The level of physical phenomena, in turn, suggests a much stronger sense of physicalism one is used to encounter in the philosophy of the mind. ”Physicalism” is to be interpreted here as requiring that the mental is not only explainable in physical terms but that possible explanations must also use physical laws. However, at the present stage of our knowledge it is impossible to imagine that conscious acts might be someday described in terms of actual physical laws. The level of neural processing is no less problematic. The neural sciences do provide us constantly with new information about the neural basis of different brain functions, but at this level the philosophic counterarguments really start to take effect. As I have already pointed out on several occasions, the majority of researchers seem to think that a neuroscientific theory of the mind is bound to eliminate the

qualitative features of consciousness along with intentionality and subjectivity. Hence it is not a very favourable view.Yet again if the first two levels of description are taken into consideration, one might be able to provoke an ontologically less radical view but one would not be able to avoid the troubles with language and meanings. In this latter case, one would also have to encounter the problems that come with every-day psychology or ”folk psychology” as it is known in philosophical literature.

The outcome of this consideration is that none of the levels of description provides a satisfying point of departure for an accurate theory of consciousness. If one focuses on the lower levels of organization, the essence of consciousness is lost. On the other hand, if the higher levels of organization are taken into consideration, one is faced with insoluble philosophical problems. What we have in our hands is evidently a no-win situation, which is a direct result of the traditional conception of levels of description that is included in the discipline of cognitive science. Others have not failed to see this defect. There is an increasing awareness that the traditional cognitive and neural levels of explanation might not be the correct ones and that, hence, a correspondence between them has so far not been reached (see e.g. Searle 1992; Bechtel 1994; Revonsuo 1997).

This defect has also given birth to a specific branch of cognitive science known as *cognitive neuroscience*. This project is set to bring the psychological and neural levels of explanation closer to each other by concentrating on how the cognition is produced by the brain (see Kosslyn 1994; Gazzaniga 1995). Just like cognitive science cognitive neuroscience is also an interdisciplinary research method, which openly combines elements from philosophy, psychology, computer science, linguistics, and neuroscience. The main idea is to endorse the approach of neuroscientific theory according to which cognition is not thought to consist of three basic levels (semantic, syntactic, and mechanic), but that consciousness is, instead, seen as a product of a multitude of various levels of organization. In practice, the simplest levels of organization can be constituted, e.g., by single neurons and the more complex levels by neuronal populations at the cortical level. However, cognitive neuroscience has not yet been able to explain how cognition arises from the brain and which level of organization in the brain is responsible for the birth and existence of consciousness. At least that is what is usually thought.

I believe cognitive neuroscience is an impressively fruitful approach to consciousness. However, I do not agree with the notion that consciousness cannot yet be explained in terms of neural processes. Therefore, I am going to use the rest of this chapter to present a model which shows that to a certain extent mental states can be described as certain neural occurences in the brain. I am aware that this will require that the problems related to meanings, internal representations, and qualitative features of consciousness have to be met in some way. Of course, the natural answer is to examine mental states, the processing of which does not involve symbolic processing, meanings, or language –and this exactly what I am going to do. Hence, throughout this study I have concentrated (and will be concentrating) solely on auditory sensations, which have no semantic content. The reason for this has now been explained.

In fact, this is not an entirely new idea. Within the cognitive sciences there has been a growing interest in exploring non-conceptualized experience instead of logical and linguistic symbolic processes –only for different reasons. The motive behind this movement has been the widening of the scope of scrutiny to involve implicit cognitive processes such as intuition, empathy, and aesthetic appreciation (see Valentine 1997). The sudden

interest in implicit cognitive processes is, in turn, a result of the recent findings in empirical neuropsychological studies, which have shown that perception, memory or even problem-solving can occur non-attentatively; that is, outside conscious awareness (see e.g. Mäkinen et al. 1995).

I am going to dedicate the rest of this chapter to presenting a model for the processing of non–conceptualized auditory information in the human brain. The model will rely fairly heavily on the distinction between attentive and non-attentive information processing.

4. In Search of a Model for Auditory Information Processing

In an article titled ”In search of the science of consciousness” (1994) Antti Revonsuo has considered some of the main approaches and paradigms within which a credible theory of the mind might be formulated. Revonsuo poses three questions, which he thinks might form the possible cornerstones of the future science of consciousness and on which research ought to be focused. These topics could be expressed in the following way (see Revonsuo 1994, pp. 265–266):

(1.) What is the phenomenological organization in the brain which makes possible conscious experience?
(2.) How can the brain have a multimodal but still coherent sensory information, and how does it have access to it?
(3.) How can an organism adapt its behavior in respect to the experiential world and, by doing so, increase its chances of survival and of producing offspring?

The first question is basically a repetition of the old bone of contention; that is, how can it be explained that consciousness is subjective and has a qualitative aspect. The traditional answer is that it cannot be explained – at least in physical or neurobiological terms. That is why the majority of philosophers of the mind holds that consciousness is irreducible to brain processes. However, Revonsuo takes a courageous step further and insists that the brain has a phenomenal level of organization, which at the macro level can be seen to constitute conscious experience. On the other hand, this same organization can be interpreted as a mechanism or mechanisms, which causally brings about phenomenal experience. Furthermore, this mechanism or mechanisms can be described at the micro level in terms of neurobiology or neural sciences in general. Hence, a true science of consciousness ought to be aimed at the identification and explanation of that mechanism.

The second question underlines a theme that has been known as ”the binding problem”. The issue here is to provide a link between the macro and micro levels of organization: how can the mechanism responsible for phenomenal experience have access to a wide variety of somatosensory and perceptual information and form a unified, coherent consciousness out of that? In other words, after we have found and explained the mechanism that is responsible for bringing about the world-as-experienced, there also has to be an explanation of its function. That is the second crucial condition for an adequate theory of the mind.

The third question deals with something much more than just the survival of the fittest. The key issue is not how an organism can adapt its behavior through some output mode on the basis of the information it receives

from the outside world. What is of essence here is how an organism can have a useful picture or a model of the external world, which is itself as a whole outside of any single experiences. I am interpreting this question to lead to the most difficult topics consciousness research has to face; that is, issues of language, culture, and the social aspect of consciousness. Unfortunately, neural sciences in their present state of development can provide very few tools for considering these issues. Therefore, I believe these questions will be among the last puzzles consciousness research is going to crack.

Revonsuo claims in his article (1994, pp. 266–272) that there are only three approaches that might possibly be able to satisfy the three presented conditions. He adds that these candidates are also the only ones which can come to terms with the problems related to the ontological status of consciousness. After all, it is no secret that others than philosophers tend to neglect ontological issues for various reasons. Anyway, the three approaches are *the neuropsychological*, *the cognitive*, and *the neurobiological model*. All the models share some general features, but each has more or less a unique point of departure. I will try to briefly outline these views.

The invention of *the neuropsychological model* was inspired by the phenomenon of implicit knowledge. This phenomenon was encountered in neuropsychological experiments with patients whose injured brains resulted in an impaired performance of certain cognitive functions. The results were surprising. Patients were found to have an implicit knowledge of the presented stimuli, even though the brain parts necessary for the performance of the cognitive function in question were damaged (see Young 1994). For instance, patients, who had damaged some parts of their primary visual cortex and reported to be completely blind in respect to those areas of their visual field, were often able to guess the location of the stimuli. In other words, even though the patients were incapable of being conscious of the stimuli and even claimed not to see anything, they still had some awareness of them (Weiskrantz 1980;1987). The phenomenon of implicit knowledge has been found to occur also in cases of tactile sensing (Paillard et al. 1983), face-recognition (see e.g. Bauer 1984; Rizzo et al. 1987; Renault et al. 1989), and in other neuropsychological syndromes, such as amnesia (see Schacter et al. 1988; Young & Haan 1990).

The phenomenon of implicit knowledge, where a subject possesses information without knowing it or being phenomenally aware of it, was a new discovery which the existing theories of consciousness and cognition could not explain. Neuropsychologist Daniel Schacter took up the challenge and formulated a model which could come to terms with this phenomenon. The model suggested that there is a common mechanism, *Consciousness Awareness System* (CAS), which underlies all conscious experience from perceiving to thinking. The idea is that, in addition to the modular mechanisms that process, say, different aspects of stimuli, the involvement of a different type of mechanism (i.e. CAS) is required in order for the stimuli to reach consciousness. This central system interacts with different specified information processors (e.g. auditory modules), and these connections must be intact if conscious experience (e.g. auditory sensations) is to be produced. Schacter thinks that the described cases of implicit knowledge are the results of a situation where some of these specific information processors or modules are disconnected from the central conscious mechanism (CAS). (see Schacter et al. 1988; Schacter 1990; Revonsuo 1994, p.267)

The second approach, *the cognitive model*, is based on the work of Bernard J. Baars (1988;1994) in which he concentrates on comparing conscious and unconscious psychological processes. The processes have been found

to have two basic dissimilarities. Firstly, conscious processes are very inefficient, slow, and they make a lot of mistakes. Unconscious processes are, in turn, very efficient, rapid, and they make relatively few mistakes. Secondly, conscious processes can have different contents at different times, but at any one time there can be only one on-going process. On the basis of experiments on selective attention it has been suggested that this is due to the fact that the conscious system has a limited capacity: similar or even dissimilar cognitive tasks are thought to operate serially and hence compete with each other for the same processing resources causing interference. On the other hand, unconscious processes are specified, independent, relatively isolated, and operate in parallel. Therefore, they receive no interference from other unconscious processes and can maintain their efficiency. (see Revonsuo 1994, p.268)

Baars has drawn quite radical conclusions from these dissimilarities. He claims that the reason why the serially operating conscious system can process and experience only one stream of information at a time is that conscious operations are distributed *globally* in the central nervous system. By this he means that there is only one system within which the same information must be simultaneously available to all of its subsystems. This can only be achieved if there is only one specific message at any one time. Naturally, there also has to be a mechanism or a framework that makes all of this possible. According to Baars, what facilitates the conscious experience of information is a central system, *a global workspace*, within which all unconscious processes are united and the global distribution of information is carried out (see e.g. Baars 1994, pp. 155–160). (Unlike the neuropsychological model Baars' suggestion includes a postulation of the neural substrate of the global workspace in the form of "extended reticular-thalamic activation system" [ERTAS]; see Baars 1988).

At this point we can clearly see the similarities between the neuropsychological and cognitive models. Both approaches insist that there must be a central system or a mechanism at the macrolevel which unites conscious and unconscious processes into one coherent phenomenal experience. However, the third approach, *the neurobiological model*, has a somewhat different notion of what constitutes consciousness. This model does not accept the idea of a single spatially located neuronal system of the brain which brings all forms of information together and is causally responsible for the creation of phenomenal experience. Instead, it holds that consciousness is a property of distributed neural activations; that is, the source of consciousness is distributed on the micro level. (see Revonsuo 1994, p.269)

Theories that could be loosely defined as versions of the neurobiological model have been presented independently by several researchers including Damasio (1989:1990), Edelman (1989), Crick and Koch (1990), Calvin (1990), and Posner and Rothbart (1991). Although all of them share the general belief that consciousness is distributed (i.e. that brain processes which underlie phenomenal experiences are totally fragmented), opinions differ in respect of where and how the binding of information occurs. For instance, Damasio (1989) holds that the coherence of conscious experience is achieved by synchronous, time-locked activations of anatomically separate neuron groups. Edelman (1989) claims, in turn, that the unity of experience is due to the re-entrant signaling between different regions. Crick and Koch (1990) suggest that phenomenal experience is facilitated by frequency-locked oscillations of neurons. Instead of the differences related to the specific implementation of the binding, I would like to point out another similarity that can be found at least from Damasio (1989), Crick & Koch (1990), and Posner & Rothbart (1991); that is, the role of attention in conscious experience. The aforementioned

researchers have stressed that there also has to be a mechanism related to the switching of attention, which will help to explain why some representations reach the level of awareness and why others remain unconscious. We will discover later that attention does have a central role in the explanation of consciousness.

Now, the truly interesting question is which of the three models has the right idea regarding the constitution of consciousness. Before answering this it should be noted that all the models take ”constitution of consciousness” to mean the whole of phenomenal experience and not just some sensory mode of it. In this sense the scope of these approaches is too wide for our purposes. After all, our interest here is restricted solely to non-conceptualized auditory sensations. A detailed model of the auditory system was presented in chapter II, which showed that the neural coding of auditory sensations was carried out by several parallel processing subsystems. It was also made clear that the subsystems performed functions peculiar to them quite independently. What is at issue here is what part of the auditory system makes some neural representations remain unconscious and others reach consciousness and how does it achieve this? We have two possible ways to explain this phenomenon: The cognitive and neuropsychological model holds that there is a central system or mechanism which is in charge of this function. The neurobiological model, in turn, claims that conscious experience is due to distributed neural activations. Which of the two alternatives is the correct one?

In the next section I am going to present a model of auditory information processing which states that both of them is partly right. It will be shown that different spectral features of auditory stimuli are constantly processed by independent, unconscious neural processes and compared with each other in a short-term memory store. It will also be shown that there are unconscious attention-switching mechanisms, which causally bring about the situation that a deviant stimulus (deviant in respect to preceding neural representations) is brought under the attention of the conscious, capacity-limited central executive.

In other words, the cognitive and neuropsychological model are right about the fact that at least some central system is needed to facilitate phenomenal experience. However, they are wrong about the fact that this system would solely guarantee the unity and coherence of conscious experience. Auditory stimuli are fully processed preconsciously, before they reach the central executive. Thus, the central system has no integrative or unifying function. On the other hand, the neurobiological model is also correct in insisting on the distribution of consciousness: the functional specificity of the subsystems and the fact that different spectral features of even a single auditory stimulus have independent neural representations seem to support the notion that the complete contents of information are not broadcast across the whole conscious system (which in this case would be the auditory system). Consequently, Baars’s idea of a global workspace is not applicable in our test case and is probably false.

5. Auditory Sensory Memory: ”Short” and ”Long” Sensory Stores

The possession of accurate knowledge about the surrounding environment requires constant interpretation of the sensory input as well as comparison of it with previously received information. The human information-processing system has two features which make this possible: (1) the capability of selective attention, which

allows us to focus on only some stimuli at a time while ignoring others; (2) the capability of memory storage, which facilitates the retaining of stimulus information in the brain before and after we attend to it. Thus, human information processing can be said to consist of three types of closely related processes, which, analyze and synthesize the received information, store it in the memory, and bring about the phase of attending to it. For an adequate model of information processing there is a requirement to provide explanations for the specific locations and functions of these processes and, furthermore, explicate their reciprocal relations.

One of the early attempts to present a model of the human information-processing system was made by Broadbent (1958) who proposed three different memory-store structures: the sensory storage of unlimited capacity, short-term storage of limited capacity, and long-term storage. According to this ”pipeline” model, some information – selected by an attentional filter – is conveyed in a fixed serial order from sensory storage to short-term storage for recognition. From there the information can later be coded into long-term storage. The model also included two controversial claims, one being the location of a selective-attention device right after sensory storage. The second claim was the idea of information feedback loops, which allowed ”top-down” influences from higher-level information to lower-level recognition. However, empirical evidence suggests that Broadbent’s proposition is severely mistaken (see Cowan 1988). The assumption of a fixed order of stores seems unacceptable, since short-term storage is found to require prior long-term information, e.g., of pattern recognition (Bower & Hilgard 1981). The early-filter theory must also go, for it has been proven that a selective-attention mechanism for short-term memory is not necessarily needed in order for information to be coded in long-term storage (see Balota 1983). Finally, Broadbent (1984) has himself later acknowledged some additional defects in the model by claiming it characterizes the subject only as a passive recipient of information.

Current views characterize the processing of auditory information in the following way. Auditory stimuli are constituted in temporal patterns, the processing of which requires that traces of the input stimuli have to be retained in the brain – at least for a short while. This important task is performed by an auditory sensory memory, better known as ”echoic memory” (Neisser 1967), which represents physical features of the stimuli. The sensory memory is thought to contain two phases or two types of memory: a ”short sensory store” which retains unanalysed auditory traces for a duration of 150–350 ms, and a ”long sensory store” in which more processed information is preserved for 10–20 seconds (Cowan 1995).

The existence of such a distinction was first suggested by Massaro (1972:1975) and Cowan (1984), and there are good reasons to believe that the notion is correct. Even though it is known that auditory sensory information can be preserved in the brain with diminishing accuracy for about 10 seconds (see Elliot 1970; Hawkins & Presson 1986; Sams et al. 1993), there are still uncertainties as to how the decaying of sensory memory occurs. A paradigm launched by Sperling (1960) suggests that the phenomenon could be characterized as a passive process which is dependent upon the lifetime and the decay of neuronal activity. The claim is supported by Lü et. al (1992a;1992b) who have discovered that behaviorally measured memory of the loudness of a tone agrees quite well with the exponentially decaying memory trace lasting less than 10 – provided that auditory input is not interrupted.

However, psychological experiments in which the auditory stimulus is interrupted seem to contradict the presented paradigm. An example of such experiments are auditory masking studies in which the perception of a

target stimulus is changed by presenting another either more intense or longer auditory stimulus. The results of these studies indicate that the categorization of elements such as the duration or pitch of the target sound is impaired if the masking sound either precedes (backward masking) or follows (foreward masking) the target stimulus within a time-span of 250 ms (see e.g. Massaro 1972;1975; Kallman & Massaro 1979;1983). Experiments with the so called ”suffix effect” (see Morten et al. 1971) have yielded the same kinds of results: subjects are found to retain more accurately the last stimuli from a series of tones when the time delay between the suffix and the final stimuli is increased.

To summarize, the duration of a memory trace can range from only a few seconds up to 20 seconds depending on the frame of reference and the nature of experiment. This evidently conflicts with the original estimate of the duration of sensory memory which held the decaying-time to be approximately 10 seconds. Thus, a distinction between the two phases of sensory memory is needed to explain the different lifetimes of memory traces.

The short, pre-perceptual store is a process-like entity which designates the experiencing of the actual sensation persisting for 200–350 ms after stimulus onset. The persistence of sensation determines the qualitative features (e.g. loudness) of the perceived sound but it also improves discrimination by prolonging an information extraction process (Cowan 1987). Information retained in the short sensory store can only be about brief sequences of sounds provided that they fall within the temporal window of integration: it has been shown that auditory input extends over time, allowing information from the beginning of the sound to be integrated with later portions of of the sound (see e.g. Zwislocki 1960;1969; Hari 1995; Hari & Loveless 1995). However, it has been shown that early neural representations seem to be able to encode not only physical features of repetitive stimuli, but also abstract attributes corresponding to simple concepts (”rise”, ”fall”); that is, to derive a common invariant features from a set of individual varying physical features (Saarinen et al. 1992). Nevertheless, it is certain that the short storage cannot contain separate information about specific sequence components, for it integrates a sound sequence as a holistic experience of a single event. The long sensory store is, in turn, more a memory-like entity, which is able to preserve information about more complex temporal stimulus sequences lasting several seconds. Contents of the longer storage are not experienced holistically as continued sensations but rather as a vivid memory of the past stimulation (Massaro 1972).

In an information-processing model presented by Cowan (1988) the two sensory stores have the following tasks. An input stimulus first enters the sensory store, which retains its physical features for a period of up to several hundred milliseconds. At the same time information in the long-term store is activated, which in turn produces the second phase; that is, the stimulus coding and short-term storage of an activated set of codes from long-term memory. Long-term memory, thus, helps in interpreting the on-going flow of auditory input, and it is also constantly updated by new, received information. Activated codes corresponding to input stimuli to which the subject has become habituated remain in short-term store but outside consciousness. However, stimuli that deviate sufficiently from the neural representation of the prior stimulation (or are in some other way significant to the subject) may enter the focus of attention. They can make an attention call to the central executive, which controls voluntary attention (= a process during which some inputs are intentionally placed in the focus of awareness), and, hence, reach consciousness.

Cowan's (1988) model could be summarized in the following way. During the first phase of perception the stimulus activates the long-term memory network, which then converges upon a set of featural and semantic categories. In the second phase of perception information that has entered awareness is used to start a more extensive search of long-term memory. This latter search will take into account additional aspects of the context in which the stimulus occured. The further processing of relevant information is carried out in working memory (= a temporary storage of information necessary for performing complex cognitive tasks; see Baddeley 1986) under attentive control (Näätänen 1992; Cowan 1995).

6. Automatic and Controlled Information Processing in Audition

When some promising approaches to creating a credible science of consciousness were charted earlier, one of the key issues turned out to be whether consciousness is distributed or centralized. The importance of the matter was dictated by the need to come to terms with the distinction between conscious and unconscious processes. First of all, the different natures of these processes (conscious ones being slow and inclined to make a lot of mistakes, with unconscious processes being effective, rapid, and accurate) required explanation. The second and probably more significant issue was to determine why some processes reach awareness and others remain unconscious in the first place. The differences were thought to be due either to neuronal activation or to a central mechanism of limited capacity, which did not only make possible consciousness but also participated in the process that determined which processes were put under attentive control. The discussion of sensory memory and of Cowan's (1988) model presented above is filled with the same thematics. Similar distinctions that occupy the minds of cognitive scientists seem to have a central role also in studies concerning auditory information processing (e.g. "pre-perceptualized/conscious", "pre-attentive processing/processing under attentive control"). Furthermore, usage of such concepts as "central executive", "selective attention", "attentional filter", "attention call" and so on make direct references to theories of a central conscious system proposed by the cognitive and the neuropsychological models. Due to the interdisciplinary nature of this study and to the wideness of the vocabulary used in different fields of research I believe that an explication of some of the key concepts is in order. Therefore, I will devote this section to providing a brief clarification of the basic terms on which the auditory information processing-system presented and proposed in the last section of this chapter is built.

The two different types of processes to which I have so far refered with the rather vague terms "conscious" and "unconscious" can be better explained with the concept of automaticy. Although the precise meaning of the concept might also vary depending on the frame of reference and scholar, the following definitions help to explicate the differencies between the two processes. Posner and Snyder (1975) have suggested that a process is "automatic" if it occurs unintentionally, preconsciously, and without producing interference to other ongoing mental activities nor receiving interference from other concurrent processes. Schneider and Schiffrin (1977), in turn, proposed that the opposite of automatic processing is "controlled" processing, which is conscious, intentional, and capacity limited. Thus, "unconscious" processes can be seen as instances of automatic processing which is rapid, efficient, parallel (i.e. receives no interference from other ongoing processes) and not limited by short-term memory capacity. Correspondingly, "conscious" processes belong to the mode of controlled

processing, which is serial, slow, effort inducing and capacity limited (and hence prone to receive interference from other concurrent processes). Unlike automatic processes controlled processes are voluntary, subject-regulated and attentional.

Of course, this distinction is not as clear-cut and unproblematic as it seems. For instance, Kahneman and Treisman have (1984) suggested that at least three levels or degrees of automaticity can be postulated: perceptual processing can be either strongly, partially, or only occasionally automatic. It is widely accepted that automatic processes can encode physical features of auditory stimuli but it is somewhat questionable whether they can also extract semantic features. There are views for (Velmans 1991) and against (Holender 1986). It has also been suggested that automatic processes might bring about an attentional switch to the processed sensory input (see Näätänen 1990) and cause either an increase or a decrease in arousal (Posner et al. 1976). Futhermore, one has to acknowledge the difference between ”automatized” processes (e.g. functions related to the driving of an experienced motorist) and those processes which are automatic by nature (e.g. neural processes found in newborns). Some of the functions of an experienced driver have automatized during several years of training. However, it has been showed that in newborns, an occasional pitch change in a repetitive auditory stimulus elicits a negativity resembling the MMN obtained in adults (Alho et al. 1990). In other words, there are some fundamental brain mechanisms which are automatic by nature and not only in nature.

Finally, besides these open issues there is still the more metaphysical question of whether information processing can even occur totally apart from consciousness.

The mode of controlled processing leads us to the issue of attention. As was previously mentioned early studies on selective attention proposed an early-filter theory, in which unattended input was blocked before perception (Broadbent 1958). However, it was soon discovered that more unattended processing took place than was previously thought (see Moray 1959). These findings led to the so called ”late-selection” theories according to which all aspects of every stimuli were fully processed (i.e. perceived) irrespective of the direction of attention. Hence the role of attention was thought to be only in the selection of some stimuli for further processing and responding (Deutsch & Deutsch 1963). In the case of auditory attention research, the filter paradigm has not lost its vitality. Though, it seems likely that some intermediate view between the early-selection and late-selection theories is the correct one (see Cowan 1988).

Now, it is important to understand that when a subject is said to be aware of or attending to some stimulus, the item is always processed by the central executive. If the central executive is not involved in the processing, the stimulus cannot be within the focus of attention and the subject cannot be consciously aware of it. By the ”central executive” or ”central processor” we simply mean controlled, limited capacity processes (Schneider & Schiffrin 1977) and not necessarily an entity-like system. It is necessary to underline this distinction, for I am sure that many are tempted to draw a parallel between central conscious systems found in cognitive science – such as Baar’s (1994) global workspace or Schacter’s (1988) Conscious Awareness System – and the central executive. However, we are dealing here only with modality-specific auditory information processing, while the aforementioned examples are meant to function as the unifiers of all conscious experience from thoughts to tactile sensing. Solely the fact that many components that participate in controlled conscious auditory information processing (e.g. working memory; see Baddeley 1986) are supposed to be formed from

combinations between the central executive, sensory memory stores, and other subcomponents is sufficient to make the point. Therefore, the questions of whether consciousness is centralized and whether the auditory information-processing system has one or more central units should not be mixed. After all, it could be that the issues have no bearing on each other.

7. An ERP Component Called *Mismatch Negativity* (MMN) Reflects Change Detection

Event-related brain potentials (ERPs) are tiny electrical brain responses that are usually caused by and time-locked to external stimuli but which can also be generated in association with the occurence of internal events within the brain. They can be investigated by *electroencephalogram* (EEG) in which they appear as small changes normally obscured by larger spontaneous brain waves and rhythms. However, by summating and averaging brief EEG epochs over many presentations of the same stimulus, the electrical activity related to the neural processing of the stimulus under investigation can be revealed with millisecond accuracy (Picton et al. 1983). After the averaging procedure, the EEG activity of the target stimuli is enhanced and the randomly occuring spontaneous waves are reduced leaving several distinct ERP deflections (i.e. waveforms) for study. The earliest ones are generated in the brainstem and they occur within the first 10 ms from the onset of the stimulus (Picton et al. 1981). The brainstem responses are followed by middle-latency responses detectable 15–40 ms after stimulus presentation. Their origin is in the auditory cortex (Celesia 1976), and they are mainly determined by the stimulus features. The late responses, in turn, consist of a large wave complex N1-P2 with peak latencies between 50–200 ms and of some other later components, such as a positive wave P300 peaking around 300 ms (see Pritchard 1981). Many of the late deflections are thought to reflect stimulus processing related to memory and attention (see Näätänen 1992).

A new ERP component called *mismatch negativity* (MMN), peaking at 100–200 ms from stimulus onset, was found by Näätänen et al. (1978). The MMN is elicited in response to infrequent change in any repetitive stimulus such as frequency (Sams et al. 1985), intensity (Näätänen et al. 1989), spatial location (Paavilainen et al. 1989), or duration (Näätänen 1990). It is generated in the auditory cortex (Hari et al. 1984; Sams et al. 1991; Alho et al. 1993) and partly in the frontal lobe (Giard et al. 1990).

The elicitation of the MMN can basically be explained in two alternative ways: a) it can either be seen to be generated by new afferent neural elements, which correspond, e.g., to the frequency of the deviant stimulus but do not participate in the processing of the standard stimuli, or b) it can be interpreted to reflect the neural processing of the stimulus difference or change (Näätänen 1984). However, there is strong evidence that the first alternative is incorrect; that is, that the MMN cannot be explained in terms of the ”fresh” afferent neural elements (see Näätänen et al. 1987; Näätänen, Jiang, Lavikainen et al. 1993). Instead, the elicitation of the MMN is explained by a mechanism that encodes repetitive stimulus features into short-lived neural representations in sensory memory and compares them with auditory input: if the ”deviant” stimulus is found to differ from the neural representation of the standard stimulus, the discrimination process generates the MMN (see Näätänen et al. 1978; Alho et al. 1986; Näätänen 1995).

Thus, the MMN reveals the mechanism of the auditory sensory memory in the human brain: it implicates the existence of memory traces belonging to the echoic memory, which can represent the physical features of the repetitive stimulus accurately (Näätänen, Paavilainen, Alho et al. 1989; Näätänen, Paavilainen, & Reinikainen 1989) but also contains precise temporal information about the sounds (Nordby et al. 1988a; Schröger et al. 1992; Näätänen, Jiang, Lavikainen et al. 1993). It has been found that the MMN model has the capacity to process more than one stimulus feature at the same time (Nordby et al. 1988b). This is facilitated by the fact that frequency, intensity, and duration are encoded in separate neural traces (Giard et al. 1995), and that there seem to be separate detectors, at least, for frequency and location of an auditory stimulus (Näätänen et al. 1988). Sensory memory traces reflected by the MMN are also capable of storing information of very complex sound structures (Tervaniemi et al. 1993). However, memory representations for simple and complex sounds seem to be located in different fields of the auditory cortex, and, furthermore, at least partially different supratemporal neuronal populations carry out the processing of changes in these sounds (Alho et al. 1996). Studies have also shown general right hemisphere preponderance in sound processing and change detection (Paavilainen et al. 1991; Levänen et al. 1996).

The change detection mechanism generating the MMN is found to function independently of attention; that is, that the discrimination process is active even though the subject maybe engaged in another task and hence unable to attend the auditory stimuli (Näätänen, Paavilainen, Tiitinen et al. 1993; Paavilainen et al. 1995). For instance, it has been proven that stimulus frequency is fully processed and encoded in a neural memory trace even in the absence of attention (Paavilainen, Tiitinen, Alho, Näätänen 1993). The elicitation of the MMN was also observed during slow wave sleep (Csepé et al. 1987), though some later studies have failed to corroborate this (see Paavilainen et al 1987; Winter et al. 1995).

It is generally thought that the functional significance of the change detection mechanism is to cause involuntary attentional switches to changes which are of potential relevance to the organism. According to this view, the system monitors automatically acoustic input and may produce an attentional ”interrupt” signal when the deviation occurs (see e.g. Novak et al. 1990; Schröger 1994; 1996). The actual source of this signal is thought to be related to the frontal generators of the MMN (see Giard et al. 1990; 1991). At least the prefrontal cortex and its temporal projections are known to have a critical role in orienting to physical changes in sequences of non-attended auditory stimuli (Sams et al. 1985; Alho et al. 1994). Nevertheless, it seems that the exact location of the cerebral generator of the MMN depends on which feature of a sound is changed and whether the deviating sound is simple or complex (Alho 1995).

The elicitation of the MMN (and the switch of attention as well) does not occur if the following conditions are not met: the representation of the standard tone must be 1) well-established in memory (which means it has to be repeated several times) and 2) in a currently active state (because after the trial the memory trace becomes quickly dormant, or out of the context, which means it has to be activated again or put back into the context)(Böttcher-Gandor & Ullsperger 1992; Cowan et al. 1993). Studies have shown that the minimum stimulus duration for an efficient encoding of the critical frequency information is of the order of 20–30 ms (Paavilainen, Jiang, Lavikainen, & Näätänen 1993). There has also found to occur an ”gradual” sharpening of sensory information encoded in the memory trace: after several repetitions the representation of the standard

stimulus eventually becomes precise enough to enable the change-detector mechanism to detect a ”deviant” stimulus (see Näätänen, Schröger, Karakas et al. 1995).

Finally, I would like to remind readers that studies have shown that only auditory stimuli seem to produce significant mismatch responses. For instance, in the case of the visual modality studies the results have not been as promising or useful as those received from auditory processing studies (Nyman et al. 1990). Even though the elicitation of the MMN has been found to occur in the visual information processing system (Alho et al. 1992), it is possible that the attention-switching mechanism related to the elicitation of the MMN is modality-specific and should not be interpreted as indicating the existence of a general attentional apparatus that would cover all conscious acts.

8. Näätänen’s Model of Attention and Automaticity in Auditory Processing

All the facts that we have so far learned about auditory information processing could be summed up as follows. Every discrete auditory stimulus that enters the auditory system is subjected to a rapid and complete processing of its physical features. The processing of sensory stimulus features, performed by a permanent feature-detection system, is automatic, parallel, and preconscious. Studies with MMN have proven that physical features are fully processed independently of attention. Mismatch responses to deviations in unattended inputs cannot be elicited unless the neural traces underlying MMN generation contain complete sensory information. On the other hand, information about the sensory stimuli retained in the brain has not been found to weaken or become less accurate if attention is withdrawed from the stimuli.

The degree of automaticity involved in the preconscious processing can be questioned. However, studies with the MMN indicate that it is usually ”strongly automatic” (= the focusing or diverting of attention has no impact on the processing) or at least ”partially automatic” (= attention may facilitate the processing)(see Kahneman & Treisman 1984). If the preconscious processing was found to be any less automatic, the functions of the generator of MMN and change-detection system would have to be reinterpreted. Nevertheless, there has been no reason to think that this would be the case. In stead, the major problem with the studies regarding the underlying brain mechanisms of attention lies elsewhere. The issue is whether the focusing of attention will damp the processing of the non-attended stimuli or will it strengthen the processing of the attended stimuli. In some cases the MMN has been observed to increase when attention is switched from visual stimuli to the auditory stimuli in which a change reflected as the elicitation of the MMN occurs (Alho et al. 1992). However, the general opinion is that the attention-switch away from the auditory stimuli might rather damp the MMN (Näätänen et al. 1993).

The sensory information produced by preconscious processing is stored for a while in the form of precise and passive neuronal representations of sensory memory located in the auditory cortex, which give rise to an MMN in the presence of deviant stimuli. It should be emphasized that we are not dealing here with actual memory, but a short-duration, sensory-register type of memory modality-specific to audition. This acoustic sensory memory does not include representations for meanings, interpretations etc., for it contains only basic

information about the spectral features of sounds. Besides the spectral features, sensory memory is also found to contain information about the loudness, duration, and direction of the sound and even information about its more complex features such as serial tone patterns (Alho et al. 1996). Nevertheless, preconscious sensory processes do not produce conscious perception but only provide its informational basis.

The emergence of a conscious perception can basically be produced by two different types of processes. The first one includes the previously explicated attention-switching mechanisms related to the elicitation of the MMN. The system of sensory representations of past acoustic stimuli is in a state of continuous flux, with older traces perpetually decaying and vanishing and new ones being added as a result of the updating of new input stimuli. If a solid trace exists and a deviant stimulus occurs, the process generating the MMN occurs producing a shift of focus. The second type contains N1 generator processes which can produce attention-switches, if a) appropriate, momentarily varying sensory attention-triggering characteristics of the stimulus (e.g. onsets, offsets, and changes in a continuous stimulus) are present and b) the momentary degree of excitability of the brain mechanisms responsive to these stimulus characteristics is reached. The threshold is lower if attention is directed toward the stimulus and higher if directed away from it. The N1 mechanisms differ from the MMN mechanism only in that they are not based on memory representations of previous stimuli. However, there is still the third possibility that brief, involuntary attention switches might occur even without acoustic events of attention-capturing characteristics. These ”breakthroughs of the unattended” are probably caused when sensory representations of recent stimuli are somehow brought into contact with semantic analyzers and the higher memory systems. (for a review, see Näätänen 1990)

It should be noted that the elicitation of the MMN is not always followed by an attention-switch. There are cases in which the process generating the MMN does not bring about conscious perception or a shift of focus (Lyytinen et al. 1992). It is shown that the auditorily evoked elicitation of the MMN interferes with a simultaneous, attended visual task by prolonging reaction-times and/or causing errors in the reactions (Escera et al. 1998). However, this interference is not enough to produce a shift of focus – at least not every time the auditorily evoked elicitation of the MMN occurs.

The focus of attention is brought under environmental control by the attention-switching mechanisms. Thus, their biological significance is based on the fact that they force the organism to attend to the present environmental situation by interrupting an attentional state directed. For instance, the limited-capacity system can be made to abort a current task performance and focus on some unexpected and even threatening event of greater biological significance or urgency to the organism. Now, everyone can understand the importance of this function which is central to the organism’s capacity to survive.

Risto Näätänen (1990) has summarized the presented conclusions in the following model of auditory information processing (Figure 2). The processing of acoustic stimuli can be divided into two different modes, which might be in part parallel and based on partially different sensory analysis: *task-independent*, basic sensory analysis and *task-dependent* sensory analysis. The task-independent sensory analyse is carried out by two mechanisms: a) a permanent feature-detector system, constituted by partially subcortical neuronal mechanisms, which extracts information about physical stimulus features for percepts and sensory memory, and b) a transient-detector system which triggers conscious perception whenever a momentarily varying threshold is exceeded.

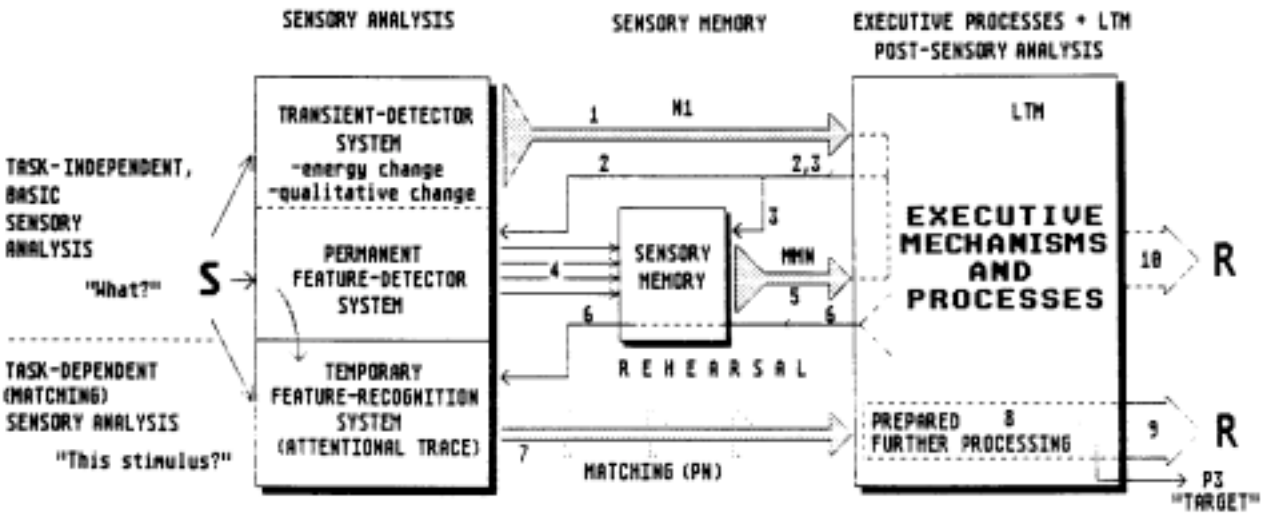


Figure 2. A model illustrating attention and automaticity in auditory information processing. Adapted from Näätänen (1990).

This latter mechanism is a N1 generator, which is activated by abrupt onsets of discrete stimuli, offsets of longer-duration stimuli, and changes in environmental energy, or even by qualitative changes in a continuous stimulus (e.g. frequency-changes) when its energy level remains constant. The transient-detector system can send interrupt signals to the central executive and cause an attentional switch either to the ongoing, conscious sensory processes or to the results of previous sensory analysis stored in the sensory memory.

The permanent feature-detector system, in turn, conveys the processed information about the physical features of input stimuli to the sensory memory, which encodes it in precise stimulus representations or neuronal traces. The traces are strengthen by repetitions of an identical stimulus. When a developed trace is solid enough, the occurence of a deviant stimulus can lead to the process generating the MMN. If this is the case, another attention-switching signal is sent to the central executive mechanism.

The task-dependent, temporary feature-recognition system is essential to the functioning of selective attention and, hence, also to self-initiated and voluntary behavior. When attention is voluntarily focused on certain relevant stimuli, the central executive mechanism maintains (supposedly with the help of the sensory memory) an attentional trace, a representation of a physical feature (or features) of the relevant stimuli, and matches or compares each input stimuli with it. If two stimuli are identical and a ”match” is found, the attentional trace ceases to exist and the target stimulus gains access to the limited-capacity system for further processing. Further processing can include either the processing of semantic stimulus features or more careful extraction of the physical features of the sound. During the time the attentional trace exists, the permanent feature-detector system continues its automatic processing in the normal way uninterrupted by the changes in attention. However, if appropriate inputs occur, the N1 and MMN generators can be activated causing involuntary attention-switches. In case these distractions occur, the attentional trace is lost for short period of time.

9. Concluding Remarks

At the very beginning of this chapter we were forced to acknowledge that modeling of information processing is troubled by a couple of rather philosophical but persistent problems related to mental representations. The first source of difficulty was argued to be the presumption that the coding of mental representations requires the existence of an internal representational system, a language of thought. However, it was made clear that the idea of a language of thought is controversial and merely at the stage of being a suggestion. For neural sciences it is more of a myth or an enigma than a concrete object of scrutiny. Secondly, it was argued that mental representations seem to involve conceptualization and reference to meanings. The problem with this issue was that there is no uncontested definition or notion as to what meanings really are. Therefore, the neuronal origin of meanings has not yet been understood, and the proposed views of the subject are quite elementary in nature.

It was suggested that both of these problems could be circumvented by setting two restrictions: (i) the study ought to concentrate on non-conceptualized auditory sensations, and (ii) the symbolic processing paradigm of cognitive science should be rejected. The first restriction erases the problems related to meanings, for representations of sounds which have no semantical content need no reference to meanings. The focusing on auditory sensations, in turn, dispenses with the conceptualization which is involved, such as in visual perception, which always falls back on meanings. The second restriction is, in fact, an indirect consequence of the first restriction: since no meanings or conceptualization are involved, there is no need for symbolic processing or for the presumption of symbolic and syntactic levels of processing. The paradigm is to be replaced with the approach of cognitive neuroscience, which claims there to be only a multitude of levels of neural organization. In other words, the question ”How does the mind represent things in this way rather than that way?” is to be reformulated in the form of ”How does the brain represent and produce auditory sensations with no semantical content?”. It is evident that such a specialized and condensed phrasing of the problem is bound to diminish the value of this study as an overall description of consciousness. Nevertheless, it provides us with better chances of reaching –if not watertight– at least more solid conclusions.

The presented model of auditory processing shows that auditory sensations are fully processed preconsciously and stored in the short-term memory as memory traces or neural representations of the auditory stimuli. It was found that deviants and discrepancies in sensory input can lead to involuntary attention calls from the automatic N1 and MMN generating processes to the central executive with limited capacity. It was also noted that the central executive can maintain an attentional trace with the help of sensory memory and voluntarily process further some selected auditory stimuli. However, this conscious processing can be interrupted by automatic attention-switches brought about by the N1 and MMN generators.

An important question related to the nature of consciousness arises from this: is consciousness distributed or centralized? Of course, the chosen approach forbids us to answer this question; it is very difficult to decide what features peculiar to the auditory system and to auditory processing might also be found from other modalities. For instance, the MMN generator seems to be modality-specific, which means that it does not have the same function in other modes of perception. On the other hand, it has been suggested that the attention-triggering processes related to the elicitation of the N1 might also occur in other modalities (see Lehtonen 1973;

Goff et al. 1977). The guess is that sensory-physiological events compete (inter- and intramodally) for the focus of attention by means of the parallel N1 generating processes leading to an attentional call for a mechanism that selects the input for conscious perception and attention (see Näätänen 1990). However, as far as audition is concerned, what is certain is that only those processes are conscious which are involved in the functioning of the central executive. So, in this sense consciousness has a unifying central system with limited capacity. On the other hand, non-conceptualized auditory stimuli are fully processed preconsciously and automatically; that is, in the absence of consciousness. Furthermore, it was shown that the auditory system consists of several independent subsystems and a multitude of levels of neural organization. They all give their specific contribution to the processing and production of sounds making the auditory information ready and complete to enter the central executive and, thus, reach awareness. So, in this latter sense consciousness is also distributed.

Chapter IV
What is Essentially Wrong with Contemporary Theories of the Mind?

The object of this study is supposed to be consciousness is general. However, I must admit that the approach I have taken here is rather ontologically oriented and dictated by the mind-body problem. Basically, my aim is to get to the bottom of two issues: i) what is the ontological status of mentality, and ii) what is the relation between consciousness and the brain. A reasonably extensive theoretical foundation for the framing of the question was presented in the first chapter; all the major positions of the modern philosophy of mind were, at least, briefly discussed. I ended up proposing Jaegwon Kim's type-identically reductive theory as the most promising starting point for formulating an adequate characterization of the constitution of mental states. According to Kim, the instantiation or implementation of a mental state *M* by a physical realizer *P* in a certain structurally-restricted system *S* can be defined as follows:

S __(P __M).

There is a good reason for believing that this definition provides a perfect starting point for further studies on the mind-body relation. Firstly, it gives a functional characterization of the mental by applying the concept of physical realization. Both notions have gained pre-eminent positions on the modern philosophy of mind and are at the heart of the current debate. Secondly, because of this fusion of modern elements Kim's definition gives us the chance to evaluate not only the theory itself but also some of the central assumptions to which the present discussion constantly refers. In other words, although it was already made clear that Kim's view has its possible pitfalls and controversial claims (such as, whether strong supervenience is the only viable option and whether it can provide biconditional bridge laws needed to carry out type-identical reductions), what is really under critical consideration here is the current debate and its generally accepted concepts and notions.

The intention of this study is to examine the prominent views of philosophy of mind from a fresh perspective by introducing a variety of solid neuroscientific facts and find out how the philosophical notions come to terms with it. As I mentioned earlier, the philosophical part of the study is limited to covering only the constitution and ontology of mental states. This means that many of the traditional sources of trouble, such as folk psychology, propositional attitudes, mental contents, meanings etc., are left outside the scope of the scrutiny. However, in order to keep the study within the set boundaries the neuroscientific part also has to be restricted to consist only of information, that supports the philosophical side, and which can be accomodated to it. Thus, the neuroscientific side of the study has concentrated solely on the production and processing of non-conceptualized auditory sensations, i.e. sounds with no sematical content. A comprehensive description of the auditory system was outlined in chapter II, which included insights on neuroanatomy and the neurophysiology of sound processing. A model of auditory processing was presented in chapter III, which showed how auditory information is processed in the brain. When the philosophic and neuroscientific parts are brought together for comparison, we have on the one side a theoretical description of the constitution or physical realization of auditory sensations and on the other side concrete evidence of how the brain produces sounds in reality.

The comparison will concentrate on testing the validity of the more basic notions of the philosophy of the mind. The constitution of mental states will be discussed in terms of the concept of physical realization. Special weight is put on the question that ought to be included as the content of a physical realizer. It will also be argued that physical realizers cannot be identified with spatio-temporally located brain states. During the comparison functional characterizations of the mental are found to be elementary and insufficient descriptions of conscious activities.

1. Brain States and the Constitution of the Mental

The current debate favors views that characterize the relation between the mental and the physical in terms of *supervenience*. It is also quite common that supervenience-relation is combined with the notion of physical realization. For instance, the relation between a mental property A and a physical property B could be described by saying that A supervenes (either strongly, weakly, or globally) on B, which is the same thing as saying that the mental property A is physically realized by the physical property B.

The relation can be postulated to hold not only between properties but also between events, predicates, sentences, facts, languages, or propositions. We could easily launch a lengthy and complicated debate on which of these accounts ought to be prefered. For instance, some think that property supervenience is fundamental, and that all other accounts can be explained in terms of it (see Kim 1984). However, I am inclined to believe that the issue of whether we should speak about neural events, states, predicates, or properties is irrelevant in the sense that it has no practical bearing on the matter. The functioning of the brain is found to be such that it causes philosophers to present quite loose definitions of their various accounts. My point is that it is not correct to assume that such terms as ”brain state” or ”brain property” refer to some stable and temporally specifiable state or property of the nervous system, but that they can also designate neural processes. Thus, too much weight should not be put on the fact that I will from now on refer mainly to property supervenience; the arguments that I am about to present will be applicable to all accounts.

Another key point is that supervenience is not thought to hold only between properties taken singly but rather between *sets* or *families* of them. Hence, we may very well imagine that mental states supervene on sets of neural properties (or on sets of neural processes). Supervenient properties also have *alternative supervenience bases*, which can either be *minimal* (= any property weaker than it cannot function as a supervenience base) or *maximal* (= a set of properties of which each might function even solely as the supervenience base). For instance, we could say that auditory sensations can either have a minimal base, which is a conjunction of the neural processes necessary for the production of sounds, or a maximal base, which contains the whole auditory system including those parts that are sufficient but not necessary for the production of sounds.

The current dispute between reductionist and antireductionist views is closely related to the idea of a supervenience or physical realization base. Hilary Putnam’s Multiple Realization Thesis attacks reductionism by claiming that it is highly likely that the same mental states are realized in different and genuine ways in different organisms. In other words, a certain pain state can have a different supervenience base in the mammalian, Martian,

and human brain. Since disjunctive predicates cannot be allowed, type-identical reductions ought to be banned. Of course, there are a number of counterarguments. For instance, Jaegwon Kim (1996) tries to circumvent Putnam's MR by postulating structural restrictions, which allow him to concentrate on the physical realization of a certain mental state only in one organism at one time. Therefore, the possibility that that very same mental state might be realized in some other way in other organisms can be excluded from the description. However, I am claiming that there are neuroscientific facts that make both antireductionist as well as reductionist formulations unsatisfying and seriously defective.

I base my claim on two indisputable facts: a) mental states cannot be solely of neuronal origin, and b) any mental state cannot have a single physical realizer. The first fact is a reflection of what we have learned from auditory processing. The auditory system consists of several independent subsystems each of which performs tasks peculiar to it. It was also suggested that different parts and neuronal populations of the auditory system show a high-degree of structural (i.e. anatomical) specifity. All this adds up to the fact that, although conscious processing and neural occurences related to the appearence of mental activity can be detected from the auditory cortex, these processes cannot take place without the contributions and the prior parallelly operating and interrelated task-performances of the subsystems. It was also argued that the group of subsystems necessary for the production sounds included the external ear, which, of course, is not a part of the brain. Hence, it was concluded that perceptions of sounds cannot have a solely neuronal origin. The only exception to the rule were the electrically caused reactivations of sensory memory. However, it ought to be remembered that the content of memory is initially processed and produced by the whole auditory system.

The second fact is an indirect consequece of the first one. Philosophers of mind have generally thought that physical realizers of most mental states are located at the cortical level, where conscious actions are found to occur. Furthermore, they have thought that the physical realizer is identifiable with a certain state of activation of a specific part of the cortex – usually discovered by empirical studies. However, what I have just stated refutes this. If one wants to designate a physical realizer for an auditory sensation, most of the auditory system has to be included in it. Though, there is the possibility that a long conjunction of neural processes is formed as the maximal supervenience base. Unfortunately, even this alternative is unacceptable. It was established that many subsystems are only sufficient in the sense that on some occasions they have *latent function* while some other subsystem carries out their task. Latent functions, in turn, cannot be included in the supervenience or physical realization base without the application of disjunctive elements. On the other hand, even if we construed a minimal physical realization base, the members of which were all necessary for the production of sounds, the situation would not be any different. It was shown that the brain has, in some occasions, a remarkable ability to recover its functions. There are cases where some of its parts are damaged, other parts of the brain which were not originally designed to or even capable of performing a certain function can take on the task of the defective part. Once more this phenomenon cannot be explained by means of the modern philosophy of the mind without reference to disjunctive elements.

The above discussion could be summarized by noting that the current theories of mind fail to meet the following three conditions for an adequate description of the constitution of mental states: i) all the independent and parallelly processing subsystems have to be somehow unified, ii) the existence of latent functions has to be

explained without a reference to disjunctive elements, and iii) the capability of the brain to recover its functions by re-organizing and replacing the task-performances of the subsystems has to be included in the physical realization base.

2. The Physical Realization of Mental States Extends over Time

Philosophical theories of the mind seem to contain – more or less implicitly – the assumption that mental states are produced by spatio-temporally located brain states. The idea that physical realizers could be identified with activations of cortical-level neurons is an example of this belief. We already established the falsity of the localizibility claim. Now, I would like to consider the latter part of the assumption; that is, the claim that physical realizations can be temporally localized.

Philosophers of the mind tend to assume that the production of mental states occurs at a certain specifiable instant of time *t*. Some views are more obscure in respect of this issue than others, but there are also quite explicit formulations. For instance, psychoneural identity theorists believe that, say, pain can be identified with C-fibre excitation. The activation of C-fibres is something that can be empirically observed, and it is, thus, justifiable to assume that we can establish the specific time *t* during which the psychoneural identity ”pain = C-fibre excitation” is thought to hold. Causal-role identity theorists, in turn, characterize mental states in terms of their roles to as mediators between inputs and outputs, which causally bring about behavioral responses and other mental states. They claim that empirical scrutiny can establish the existence of a neural correlate, ”a brain state *B*”, at a certain time *t* for a certain functionally characterized mental occurence, ”a mental state *M*”. This allows the conclusion that the identity ”a brain state *B* = a mental state *M*” holds at the time *t*. An example of a more vague notion is Donald Davidson’s Anomalous Monism, which claims that every mental event is also a physical event. I am not going to enter into Davidson’s complex and lengthy argumentation here. However, it can still be said that, since the other aspect of every mental state is always a spatio-temporally located physical event, physical realizers of mental events can be interpreted to be temporally and spatially distinguishable and specifiable neural events and processes. Probably the most explicit characterization is provided by the property-exemplification account of events, which could be said to have brought into the open what Davidson merely suggests. According to this account, an event is an exemplification of a property at a time in a physical object. For instance, if the substance *x* has the property *P* at *t*, we can say that there exists an event [*x*, *P*, *t*].

All the presented identity claims are vulnerable to the same objection that was directed against the idea of a single, cortical level realizer of mental states. However, there are also some additional points to be made, which might give further support to my claim that relations between mental and physical properties cannot be assumed to hold at one specific instance of time.

The nervous system is known to collect information for 100–200 ms before the conscious perception takes place (Hari & Loveless 1997). During this time the incoming auditory stimulus is preconsciously processed and compared with the context of priorly analyzed stimuli. If the stimulus differentiates sufficiently from the preceeding stimuli, the information is sent to the conscious central executive for further processing. In case of

sound localization, if overlapping neuronal populations are activated in rapid succession, the conscious perception could be delayed for about 500 ms (Hari et al. 1995). Despite this delay, the subjective experience of the sensation is found to occur without any significant delay. Hence, it seems that conscious sensory experience has different neuronal and subjective timing.

It has been suggested that this phenomenon could be explained by postulating the existence of a specific lemniscal projection system, which serves as a ”time-marker” for received stimuli, and that there is an automatic subjective referral of the conscious experience backwards in time to this time-marker (see Libet 1978). This sort of arrangement would, thus, make it possible that sensory experience appears subjectively to occur without any significant delay. Susan Pockett (1999) has recently provided a bit different solution to the problem. She distinguishes between ”auditory perception” (which is taken as including such functions as memory of the sensation and detection of differences between the current sensation and preceding ones) and ”auditory sensation”. The neural correlate of ”auditory sensation” is represented by the middle latency waves of the auditory evoked response, which are generated in the supratemporal primary and/or secondary auditory cortex and occur from 20 to 80 ms after the stimulus. Thus, Pockett suggests that the time lag between stimulus and sensations is not 500 ms but only 20 to 80 ms. According to her, this is approximately the same as the 50-ms time lag the experimental subjects experience subjectively.

The presented fact that physical realization does not only consist of several, parallelly operating processes, but that it also extends over time, has problematic consequences. Of course, we can say that a sensory experience occurred at such and such a time, because it is not the mental part of the description that causes troubles. If we say that an auditory sensation *A* occured at the specific time *t*, it is difficult to decide what should be counted as the physical realization or supervenience base of this sensation. There are several possibilities to choose from: if the perception is delayed, for example for about 500 ms, the physical realizer could be the preconscious, parallel processes that do the actual extraction of sensory information, either the MMN or N1 eliciting processes which send attention calls to the central executive (if either of them occured), the conscious processes of the central executive, the automatic subjective referral system of the conscious experience, or even all of them together. Naturally, auditory processing requires the contribution of all these systems. However, some of the processes overlap with each other and some of them do not. If the auditory sensation is found to be perceived at the time *t*, most of the processes have occured before the time *t* within the time-span of 500 ms.

I believe my point is quite clear. All the processes of the auditory system which contribute to the production of sounds ought to be included in the physical realization base. However, there has so far been no theory that can do this. Even if such a theory existed, it could not claim that the supervenient property and the supervenient base co-existed at a specific instance of time *t*. This is due to the fact that physical realizations of mental states extend over time. Whether or not the relation between the mental and the physical is thought to be reductionist in nature has no bearing on the matter.

3. Functional Characterizations and the Criteria of the Mental

Functional characterizations describe mental states according to their mediating causal roles between input and outputs. For instance, pain can be characterized to have the causal role of bringing about winces and groans whenever tissue damage occurs. However, auditorily evoked mental states are somewhat difficult to describe within this approach. After all, the perception of sounds does not necessarily cause any kind of behavioral outputs or even subsequent mental states. The scope of this study is limited in such a way that only non-conceptualized auditory sensations, which do not elicit any kind of responses, have been considered. Therefore, the objects of this scrutiny lack the characteristics needed for functional descriptions. How can auditorily evoked mental states be characterized then? And how can it be determined whether or not a person has an auditory sensation? I believe that these questions are interrelated in the sense, that if we can stipulate what can be counted as a mental state and what is not, then the characterization problem resolves itself.

The issue at hand is to find the criterion (or the ”mark”) of the mental. I already presented a lengthy discussion of the matter in chapter I in which *the intentionality criterion* was argued to be the most promising alternative. The more recent version of the intentionality approach generally holds that the criterion of the mental can be set in terms of *content intentionality*, which claims that mental states have the peculiar feature of containing representational contents and meanings. This notion is based on the previously explicated idea that the human brain transforms perceptual input into mental representations, which refer to the state of affairs of the external world. John Searle (1992) has made a noteworthy remark on the notion of content intentionality. Searle makes a distinction between *genuine* or*intrinsic intentionality* and *derived intentionality*. Human mental states are thought to be intrinsically intentional, or intentional in their own right. The concept of derived intentionality is, in turn, attributed to such systems as computerized intelligences, which seem to be intentional, but which only repeat intentional sentences programmed into their memory. Thus, computers cannot possess genuinely intentional states, for the contained intentionality and meanings of the programmed states are derived from the mental processes and used language of the programmers (see chapter I: 1.1.2).

The intentionality criterion is probably one of the best resolutions to the problem of marking the mental that current theories can provide. Nevertheless, it is quite powerless in respect of those auditory sensations in which we are interested. The presented criterion assumes that mental states can be recognized by their genuinely intentional contents (i.e. mental representations). However, it is not certain that non-conceptualized auditory sensations are presented in the brain as mental representations. I argued earlier that the whole idea of mental representation is so full of problems that one ought to remain on the level of neural presentation. The model of auditory processing that was presented showed that auditory stimuli are fully processed preconsciously, and that even the conscious processes of the central executive are merely neural occurences. Thus, mental representations might not be needed in explaining the perception of non-conceptualized auditory stimuli (the last chapter is dedicated to explicating this point, but for now we have to take it as granted). If this is so, then the intentionality criterion is evidently of little use.

I am proposing another criterion for auditorily evoked mental states, which might be characterized as empiristic in nature. Many of the ERP studies are conducted in the condition, where the subject’s attention is

directed away from auditory stimuli by engaging him/her in a reading task. Even in this inattentive condition, the brain processes preconsciously and constantly auditory input, and in case a stimulus deviates sufficiently enough from the prior input, an attention call is sent to the central executive by the generators of the MMN after which the subject becomes aware of the stimuli. The MMN generator is related to orienting functions; that is, diverting the organism's attention to potentially significant stimulus changes in the environment. The auditory system is known to have such a change detection system, but there is no absolute certainty as to what this exact mechanism is. However, it has been found that the architectonics of the system generating the MMN is suited to serve the attention-trigger function. The general opinion among brain researchers holds it to be such (see, e.g. Schröger 1996). Hence, the elicitation of the MMN (at least in the reading condition) might function as an indicator of the appearance of auditorily evoked mental states. Since the generator of the MMN and the conscious central executive are only sets of neuronal processes, it seems that *some* auditory sensations can be described at the level of neural representation. Of course, one has to keep in mind that the elicitation of the MMN is not always followed by conscious perception or a shift of focus. Therefore, I am merely adressing that there are some cases in which the elicitation of the MMN can function as an indicator of or a criterion for mental states.

4. Summary

When the presented set of neuroscientific facts and some of the elementary conceptions of the philosophy of the mind were compared, the results did not flatter contemporary theories of the mind. It seems that whatever the philosophers of the mind are studying it is certainly not human consciousness or the human brain. Their disregard for the development of neural sciences has led to the philosophy of the mind being left behind. This attitude is rather surprising, for other consciousness researchers have enthusiastically and gladly accepted brain research as an inseparable part of forming theories.

Although the presented comparison reveals the pitiful state of the contemporary philosophy of the mind, it also gives some useful ideas as to how the constitution of mental states ought to be explained. The concept of physical realization is found to have three major defects. Firstly, in the current form it cannot come to terms with the fact that the human brain operates in parallel. Secondly, it is inconsistent with multiple realization which is found to occur constantly in the brain. Thirdly, the concept cannot explain that physical realization of mental states extends over time, and that the actual sensations can be delayed for several hundred milliseconds – even when it is experienced as occuring instantly. It is also argued that there might be an empirical method for detecting the occurence of a mental state. In other words, it could be the case that the criterion of the mental is located in the brain.

In the next chapter I will take another look at the constitution of mental states. The presented criticisms and suggested improvements will be taken into consideration. A dispositional theory for the constitution of mental states will be formulated. I will also argue that mental states – including their qualitative aspects – can be explained in purely neuronal terms. Thus, consciousness is suggested to be nothing more than brain activity.

Chapter V
A Homunctionalist Interpretation of the Auditory System: Mental States as Abstract Dispositions

I am sure that an informed reader might feel a little disappointed at this point. I have all along promised to present something radically new, but now that we have reached the finale of this thesis the title of the chapter indicates only a repetition of old ideas, such as homuncular functionalism and dispositions. However, what I am about to present in the following pages has a significant sense of freshness about it. I will rely on conventional and widely accepted ideas and conceptions related to functionalism and dispositions, but my intention is to reconstruct a genuinely new kind of combination from them. We have so far wandered through concepts of the philosophy of the mind, neural sciences, and cognitive science in a way that might have left the reader puzzled by the purpose of it all. Now it is time to summarize all the previous discussions and acquired knowledge and synthesize it into a theory of the constitution of mental states.

I have advocated Jaegwon Kim's type-identically reductive theory of the mind as the most promising alternative of the current views. Kim's theory combines the notions of physical realization and supervenience, and it can be formulated as follows:

S _ (P _ M).

A group of arguments was presented, which led to the favouring of Kim's view. The current views prefer to characterize the relation between the mental and the physical in terms of supervenience. This, I argued, was due to the fact that the concept of emergence failed to make psychophysical connections tight enough and give an explanation of the relationship itself. Both emergence and supervenience are, though, quite similar as forms of nonreductionist materialism, but supervenience possesses better logical tools for explicating the mind-body relation. Nevertheless, even supervenience was found to have three major difficulties. Firstly, the notion that supervenient properties are completely determined by their base properties seems to exclude the possibility of mental causation. Hence, supervenient formulations can at best propose epiphenomenalism. Secondly, it was argued that only the "strong" version of supervenience is capable of providing the required psychophysical correlations. However, strong supervenience was shown to imply such strong nomological necessities, that it facilitates the formulation of biconditional bridge laws, which, in turn, allowed local, type-identical reductions. Thirdly, it was found that the concept of supervenience does not explicate how the supervening properties are constituted; for instance, whether are they causally produced or whether they co-exist in parallel.

Also Kim's theory was argued to have its defects. For instance, even though Kim circumvents the Multiple Realization Thesis in the Putnamian sense by endorsing only local reductions and by restricting the characterization to cover a specific type of organism (e.g. the human brain) at one time, it does not managed to come to terms with multiple realization within an organism. The origin of qualitative aspects of consciousness (e.g. subjectivity and qualia) is also left unexplained. Moreover, Kim's theory was found to suffer from all the same compatibility problems related to the presented set of neuroscientific facts, which trouble the majority of contemporary views on consciousness. These sources of trouble were listed to be the spatio-temporal localization

of physical realization bases, the setting of the criteria of the mental, and the reference to mental representations in explaining information processing.

I argued that these difficulties can be resolved by adapting the approach of cognitive neuroscience, according to which the brain perceived as comprising a multitude of neuronal levels of organization. The aim of cognitive neuroscience was mentioned to be the explication of how consciousness gets produced within this organization. A description of the auditory system and a model of auditory processing was presented, which showed that non-conceptualized, auditorily evoked mental states can be explained in purely neuronal terms. It was also noted that no current theory can provide a philosophical basis for the constitution of these mental states.

This defect evidently needs to be corrected, and this is precisely what I am about to do here. I am going to formulate a theory of the mind, which can circumvent all the presented problems and shortcomings in a way that is in accordance with the presented neuroscientific facts. The suggested theory will include a totally new way of explaining subjectivity and qualia. It is based on some general and widely accepted notions of dispositions, and it takes the form of homuncular functionalism.

1. The Concept of Disposition

Before we move on to consider the nature of dispositions, a few conceptual clarifications are in order. Convention distinguishes between two types of dispositions, *psychological* and *philosopher's dispositions*. The characteristic feature of psychological dispositions is that they are, by their very nature, possessed only by entities with minds. For instance, psychological dispositions are constantly used in our ordinary language, when someone is said to be easily irritated (= to possess the disposition of irritability) or subject to sudden changes in mood (= to possess the disposition of moodiness). Another class of psychological dispositions can be found from the philosophy of mind in which dispositional analysis is used to characterize mental phenomena. According to Gilbert Ryle's (1949) classic paradigm, to say that *x* undergoes the mental occurrence of understanding algebra is nothing more than to say that *x* is disposed to find correct solutions to algebraic equations when they are presented to *x*. Causal-role identity theorists, such as D.M. Armstrong (1968b), makes use of this notion by stating that mental states are *pure* dispositions which causally bring about certain outputs in the presence of certain inputs. The difference between Ryle's and Armstrong's accounts is that Ryle does not think dispositions are related to properties (and hence to be observable occurrences or states) but only to events standing in contingent relations. Armstrong, in turn, believes that it is precisely a property of the subject (i.e. a brain state) with the right causal powers that makes a dispositional ascription true. Although psychological dispositions are usually associated with talk of consciousness and mental states, they are not the object of this study. Instead, I am focusing on philosopher's dispositions – such as fragility, solubility, hardness, and the like – which are typically possessed by inanimate objects. The reason for this choice of approach will become apparent as I continue.

Besides concentrating on philosopher's dispositions, I am going to propose a *realist* theory of dispositions. Realism with respect to dispositions is currently the most popular and widely accepted view, which

holds that a system or an entity has a disposition due to its specific base (for the benefits of the realist approach, see Mumford 1998, pp. 37–63). Thus, *phenomenalist* accounts (e.g., Ryle's view), which see dispositions merely as the holding of certain conditionals and which deny the thesis that dispositions must have bases, are categorically rejected (for objections to phenomenalism, see Prior 1985, pp. 29–42). It should be noted that the realist analysis of dispositions does not necessarily require an explication of the structure and composition of the base. A systems can be said to possess a certain disposition without explaining what it in its base facilitates the manifestation of the disposition. What is required is the assumption that such a base exists, although it is not known what the base is and how it is related to the system's capacity or disposition to behave in acertain way. However, it is not particularly useful to know that a system possesses such and such a disposition, if the reason for having it is not explained. Thus, from the viewpoint of the realist approach there are two key questions that need answering: (1) How does a base explain the manifestation of a disposition? (2) What is the relation between a base and a disposition? These issues will be taken into consideration here.

Robert Cummins presents in his book *The Nature of Psychological Explanation* (1983) a set of excellent tools for coping with dispositional explanations. Cummins distinguishes between two types of scientific explanations. Firstly, we can concentrate on *explaining events or changes within a system*. For instance, we can ask why did a certain system S move from a state s_1 to a subsequent state s_2? In the case of dispositions, two factors determine the answer: a) why did the system acquire the disposition in question, and b) what brought about the manifestation of the disposition? The first question can be answered by discovering the causal history of the system. The second one, in turn, can be resolved by tracking down the chain of events that led to the manifestation of the disposition. In the latter case, the characterization of some set of conditions and the presence of a triggering cause is needed. Cummins calls these sorts of explanations, which express law-like and causal behavioral dispositions of systems or entities, *transition theories* (Cummins 1983, pp. 14–15). The second type of scientific explanation is related to *properties*. For instance, we can ask what is it for a certain system S to possess a property P, or for what reason does S has the property P? Now the difference is quite obvious: what is of interest here is not the change but the property itself. Cummins calls these sorts of theories *property theories* (ibid.).

As I mentioned, realists hold that an entity has a disposition due to its base, and that the explication of a disposition requires that its manifestation can be explained in terms of the properties and structure of the base. With the help of property theories the explication can be carried out. Of course, it is self-evident that dispositions cannot be explained solely in terms of their bases. For instance, if a glass breaks when dropped it does not mean that it is going to do the same thing on the moon which has only 1/6 of the earth's gravity. On the other hand, some materials become significantly more fragile at very low temperatures. Therefore, dispositional explanations have to always include triggering causes, environmental and other conditions. However, from now on these factors will be taken for granted, and the discussion will concentrate on dispositional bases.

Cummins tries to explain the manifestation of a disposition by applying an "analytic strategy" to a system. According to this strategy, a behavioral disposition or a capacity is explained in terms of the properties of the system's components and their organization. To put it briefly, the analysis of the component structure

facilitates the description of the instantiation of a (dispositional) property in a certain system. Cummins makes a distinction between two forms or methods of analytic strategy: *compositional* and *functional analysis*.

In *compositional analysis*, a system is deconstructed into several subsystems or components, which together are thought to constitute the whole system. For instance, if we want to know why some substance has the disposition of being water-soluble, all we have to do is to find out what particles constitute the molecular structure of the substance in question and in what way they bind to molecules. Therefore, we can say that the disposition of being soluble in water is instantiated in the specific molecular structure of the substance (Cummins 1983, pp. 18–19).

Compositional analysis suffices to explain only very simple dispositions. When the explained dispositions get more complex, the organization of the components also starts to matter. In these cases, compositional analysis must be supplemented with *functional analysis* in which the occurrence of a target disposition is analysed as the manifestation of a group of much simpler dispositions. For instance, we can say that a system *S* has a disposition *D* which gets manifested when a set of simpler dispositions $d_1,\ldots, d_n$ (possessed by *S*) occur. Cummins' own example of functional analysis is formulated in terms of an assembly-line, which consists of several independent and simple components. Each component has a special task assigned to it, and its function is to perform that task. The assembly-line as a whole can be explained to have a capacity to produce a certain product only because its components perform their functions as programmed and they are organized in a specific way, which supports and coheres with the whole program of the assembly-line (see Cummins 1975). In the same way biological organisms can be explained as having a certain capacity, when they are deconstructed into several independent subsystems (Cummins 1983, p.29).

We were set to find out how a base can explain the manifestation of a disposition. Now we know, that when environmental factors and other conditions are taken as given, there are two methods by which analysis can be carried out. In the most simple cases, only compositional analysis, which concentrates on the structure of the system and which explains the capacity of the whole in terms of the capacities of the components, is sufficient. When dispositions get more complex, it has to be complemented with a "top-down", functional analysis, which takes the capacity of the whole system into account and deconstructs it into several subsystems. Compositional and functional analysis do not always cohere as neatly as described. However, the most fruitful results are usually reached, when a system is deconstructed by functional analysis into several subsystems, and when the capacities of these subsystems can be localized in specific structural parts of the whole by compositional analysis.

Since we have acquired an appropriate means of explaining dispositions in terms of their bases, it is time to move on to the second question, which is related to the exact nature of the relationship between a base and a disposition. Actually, the issue of the relation between a disposition and a base contains two bones of contention. The first is the location of the dispositional bases, and the other is the nature of a dispositional base. There are basically two noteworthy views on the issue of the location of a dispositional base. Firstly, one can present several variations of the idea that the basis of a disposition is a relational property of the item possessing the disposition (see e.g. Mackie 1977; Smith 1977). Secondly, one can take the position proposed by Armstrong (1968b) and Prior (1985), which claims that the basis of a disposition is a non-relational property or a property-complex of the item possessing the disposition. I am inclined to support the latter view for reasons that cannot be

considered here (see Prior 1985, pp. 43–58). The reason why I am willing to pass over this issue quickly over is that there is a much more elementary problem related to dispositions than the dispute over whether dispositional bases should include relational properties (e.g. environmental circumstances etc.) or not. What is of interest here is the question of the precise nature of dispositional bases.

The distinction between categorical and dispositional properties concerns the debate on dispositions in two different ways. First of all, we may repeat the commonly accepted view according to which dispositional properties can be distinguished from categorical properties because dispositional ascription sentences possess a relationship to certain subjunctive (or counterfactual) conditionals not possessed by categorical ascription sentences. This is a fairly self-evident distinction, for the essence of being a dispositional property is to manifest the property in question in the presence of appropriate circumstances and an initiating cause.

However, the distinction becomes of much greater importance and also more controversial when associated with dispositional bases: are dispositional bases categorical or dispositional properties? I tend to support Elisabeth Prior (1985, pp. 62–63) on this matter and claim that dispositional bases are categorical properties. Although there are opposite views (see Mellor 1974), I believe that the idea of a categorical base is the most sound alternative. For instance, if it is said that an object has a disposition to shatter when hit with a certain force, what is argued is that that object has a certain atomic composition, which gives it certain dispositional, macro-level properties (e.g., hardness, fragility, etc.) in the presence of certain circumstances. If it is said that the base of the disposition to shatter is another dispositional property, one is faced with an unfortunate consequence. Since the atomic composition is only a dispositional property and according to realist analysis every disposition has to have a base, we have start to look for the (subatomic) base property of the atomic composition.

I am sure that by now everyone can see that we are heading towards an infinite regress. Even if this was not the case, human knowledge has its well-known limits when it comes to micro-level phenomena. Therefore, the idea of dispositional base properties does not seem very appealing. It is more useful to suppose that dispositions have categorical bases, for they can provide quite adequate and sufficient explanations for the dispositions that are encountered in the macro world. Science requires nothing more. Whether or not the notion of a categorical base is ontologically correct is something I am willing to let philosophers contemplate. It will ultimately have no impact on (and it will not bring anything substantially new to) dispositional explanations.

It has so far been claimed that dispositions must have bases, and that the basis of a disposition is a categorical, non-relational property (or property-complex). The issue of the exact nature of the relationship between a disposition and its base has still not been completely covered. What is obviously needed is some sort of dependency relation to connect the manifested disposition with the categorical base. Although I started this discussion about dispositions by stressing that psychological dispositions are not of interest here, it seems that we are ironically in a similar situation as if it were the mind-body relation under consideration. I argued in the previous chapters that strong supervenience is the only viable option for connecting the mental to the physical, and, therefore, it is so in the case of dispositions and their bases. The only problem with this line of thought is that strong supervenience was shown to entail reductionism. Thus, the counteraguments that seem to overwhelm mind-body reductions are just as effective with respect to dispositions.

The idea that dispositions may be reduced to or identified with their categorical bases can be rejected for numerous reasons. For instance, the same disposition can be brought about by several independent bases. There are many substances that are soluble but which have totally different physical composition. On the other hand, it is possible that an entity has multiple categorical bases for a certain disposition (see Mackie 1973). There are also cases in which the same base is responsible for the manifestation of several different dispositions (Harre & Madden 1975). Finally, the base can exist in the presence of appropriate circumstances and a triggering cause without manifesting the disposition. This is due to the fact that the base can contain some properties that overwhelm its capacity to manifest a disposition (Prior et al. 1982, p.253). We have already learned that the only way these phenomena can be explained is by applying disjunctive properties. However, it was made exhaustively clear that disjunctive properties cannot be allowed. Therefore, any sort of identifications or reductions are out of the question.

My intention has all along been to take philosopher's dispositions and apply them to the relationship between the mental and the physical in a way that would circumvent the mind-body problem. Hence, the fact that the discussion has ended up in the same swamp from which I initially tried to escape is particularly embarrassing. Nevertheless, I am convinced that there is no cause for desparation yet. The reductionist/ antireductionist distinction is not the only frame within which the relationship between a categorical base and a disposition has to be formulated. I am going to later present a third alternative, which will give a functionalist interpretation of dispositions. Only then will the benefits of philosopher's dispositions in respect to the mind-body problem become apparent.

2. Homuncular Functionalism

According to W.G. Lycan (1987a), *homuncular functionalism* or *homunctionalism* originates from the work of F. Attneave and it has been further developed by D.C. Dennett. Though, it was Lycan himself who formulated the first explicit characterization of homunctionalism as a part of functionalism. Homuncular functionalism has received rather reserved responses mainly due to its classic use in the philosophy of psychology. In this early sense, a component of the mind – a homunculus – was endowed with some mental ability it was supposed to explain in the first place. The reason for doing so was simply that genuine or even better explanations were not available. Evidently, this sort of strategy cannot end up in anything other than regression. However, modern homuncular functionalism in which we are here interested does not stoop to such cheap explanations. According to modern versions, homunculi are thought to explain some higher-level ability, but they are themselves seen to belong to a lower level and to possess significantly less sophisticated abilities than the ones they are suppose to explain.

Thus, modern homuncular functionalism seems a very credible program, for it presupposes a continuity in the levels in nature which, in turn, enables it to avoid circular descriptions. Dennett has provided a neat exemplification of the idea by applying it to the area of AI research:

> ...first and highest level of design breaks the computer down into subsystems, each of which is given intentionally characterised tasks; he composes a flow chart of evaluators, rememberers, discriminators, overseers and the like. These are *homunculi* with a vengeance, the highest level design breaks the computer down into a committee or army of intelligent homunculi with purposes, information and strategies. Each homunculus in turn is analysed into smaller homunculi, but more important into less clever homunculi. When the level is reached where the homunculi are no more than adders and subtractors, by the time they need only the intelligence to pick the larger of two numbers when directed to, they have been reduced to functionares "who can be replaced by a machine". The aid to comprehension of anthropomorphising the elements just about lapses at this point, and a mechanistic view of the proceedings becomes workable and comprehensible. (Dennett 1975, p.80)

Although Dennett's example is about computerized intelligence, it is hard not to notice the benefits of the approach used. Firstly, the idea of functionally specified, independent, and unintelligent subsytems is a very workable one. The previously presented description of the auditory system provided convincing evidence of the fact that this is precisely the way in which the human brain is organized.

Secondly, the reasons why philosopher's dispositions might prove to be quite useful in explaining consciousness have started to become evident. Dennett assumes that some subsystems are intelligent and that even in the simplest cases they add up to adders and subtractors. However, it was shown that in the case of audition the processing of non-conceptualized sensations does not require any kind of intelligence at all. Furthermore, it was also shown that the capacities of the subsystems do not even amount to adders: physical composition and structure determines the functions of subsystems, which can, thus, be explained mechanistically. Therefore, it seems almost natural to hold that the performed functions of the subsystems are actually manifestations of the (philosopher's) dispositions they possess. If we think only of Cummins' proposition to use compositional and functional analysis in dispositional explanations, it is quite obvious that these very forms of analysis in fact strongly imply homuncular functionalism.

The rest of the chapter is dedicated to presenting a theory of the constitution of mental states, which relies both on homunctionalism and philosopher's dispositions. Also a solution to the problem of the relationship between a disposition and its base is offered. I will end the discussion by considering some of the most obvious objections to the presented theory.

3. Mental States as Functionally Interpreted Abstract Dispositions

I argued earlier that the relationship between a disposition and its base is troubled by similar problems that can be found with the mind-body relation. Dispositions cannot be identified with their bases, but also any kind of token identity accounts are just as unacceptable. There is, though, a third alternative; that is, to adapt a view which proposes functionalism about dispositions. This idea has been endorsed at least by Elizabeth Prior (1985), John Cambell and Robert Pargetter (1986), and Stephen Mumford (1998). Functionalism about dispositions holds that different items which possess a particular disposition share some common functional essence. For instance, we may say that all the different fragile objects possess a property (or property complex) which plays the same causal role in bringing about a manifestation of the fragility when the fragile item is subjected to a suitable

initiating cause in the presence of suitable standing conditions (Prior 1985, p.83). The benefit of the approach is that there is no need to postulate some (preferably supervenient) property-to-property relation between the base and the disposition. The functional definition assumes no more than that an entity possess a particular physical property which plays a certain causal role in a certain situation.

Functionalism about dispositions does have a few problems. To say that an entity has a disposition is the same thing as to say that that entity has a certain functional essence, which means a second-order property. In the case of mental states, this view implies the traditional functional characterization of the mental. For instance, if we say that mental states are dispositions, we are actually saying that the brain possesses a property which causally brings about certain behavioral outputs and other mental states in suitable circumstances. In this case, mental states are interpreted to be psychological dispositions. There are two reasons why this view is incorrect. First of all, it was established that functional characterizations of the mental are inadequate and incompatible with the actual brain events that occur during conscious experience. Secondly, if mental states are interpreted as second-order properties, the definition requires that there are always behavioral responses or other mental states which can causally be brought about. However, the interest of this study is restricted to non-conceptualized auditory sensations, which do not necessarily elicit any kind of behavioral responses nor subsequent mental states. In fact, they do not even require the existence of other minds, social community, or language. Therefore, psychological dispositions are of no use in explaining the constitution of mental states.

However, I am convinced that the constitution of mental states can be explained in terms of philosopher's dispositions. Let's say that mental states are dispositions possessed by the brain and take as an example non-conceptualized auditory sensations. In this case, we can say that non-conceptualized auditory sensations are dispositions possessed by the auditory system. I will leave the question of the exact nature of these disposition presently open and return to it later. It was shown earlier that the auditory system consists of several independent subsystems which are both structurally and functionally specified. Hence, the application of compositional and functional analysis as well as homuncular functionalism is well justified. By applying homunctionalism we may say that the auditory system S consists of several independent subsystems $s_1,\ldots, s_n$. The dispositional interpretation in turn states that the auditory system S has a disposition D to produce or realize a certain non-conceptualized auditory sensation. Furthermore, by applying compositional and functional analysis we may also conclude that the manifestation of the disposition D occurs only when the simpler philosopher's dispositions $d_1,\ldots, d_n$ are realized in the programmed way. Naturally, the dispositions $d_1,\ldots, d_n$ are assumed to be the specified capacities performed by the subsystems $s_1,\ldots, s_n$.

When a subsystem s_1 is said to possess a disposition d_1, what is actually meant is only that s_1 has a certain physical composition (i.e. possesses a certain physical property) which brings about particular outputs in case of particular inputs. Let's take the tympanic membrane, an important part of the auditory system, as an example. The tympanic membrane has a physical composition which gives it the capacity to vibrate and convey the vibration to the ossicles of the middle ear in case air-pressure waves happen to arrive from the external ear canal. Although the tympanic membrane might seem a fairly insignificant subpart of the auditory system, the damaging of it will impair or even prevent sound processing. The physical composition of the tympanic membrane also gives it other capacities than simply the ability to vibrate. For instance, it can be easily pierced,

but it is also flexible to a certain extent. Therefore, subsystems inevitably possess multi-tracked dispositions the manifestation of which is dependent upon the received inputs and circumstances. The tympanic membrane possesses at least the dispositions of vibrability, flexibility, and piercability. This can be explicated the following way (see Prior 1985, p.97):

> A tympanic membrane t_1 has multi-track disposition $d_{mt} = t_1$ has some property which given I_1 brings about O_1, given I_2 brings about O_2, and given I_3 brings about O_3.

It is important to remember that even though the subsystems possess multi-track dispositions only one or few of the dispositions are relevant with respect to auditory processing. It should also be remembered that only one of the dispositions gets manifested at any one time depending upon the received input.

An astute reader might have already noticed that even this dispositional account is not immune to the problem of multiple realization. This is due to the fact that the definition "D = $d_1,\ldots, d_n$" seems to suggest that the dispostion *D* gets manifested only if *all* of the simpler dispositions $d_1,\ldots, d_n$ are realized in the programmed way. However, we have already learned that this is not the way the auditory system works. Not every part of the auditory system is necessary for sound production on every occasion of auditory processing. It was demonstrated that a high-degree of structural and functional specifity was found throughout the entire auditory system. This does not simply mean that the subsystems have their specified functions. For instance, it was shown that there exist neuronal populations with amplitopic and tonotopic organizations. In other words, some subsystems are sensitive to react only to some frequencies or amplitudes. If the input is out of their frequency-range, the dispositions which are peculiar to them and relevant to sound processing do not get manifested. On the other hand, many subsystems are capable of carrying out the same tasks, or in the case of brain damage subsystems can take over the functions of the defective subsystem or subsystems. Nevertheless, the dispositional account can easily come to terms with such phenomena. If the nature of the input is such that it does not activate a certain subsystem, that subsystem has on that occasion *a latent function*. After all, for a subsystem to perform a certain task is to manifest a certain disposition. If the case is such that a suitable input and suitable circumstances are not present, the disposition is not manifested. Since dispositions are not properties (but functions that are causally brought about by some property of an item when the item in question is subjected to a suitable initiating cause in the presence of suitable conditions), there is no need to postulate the existence of unacceptable disjunctive properties. What exists are only *active* or *latent* functions, which are determined by the input and conditions on each occasion. Therefore, every task-performance of every subpart of the auditory system can be included in the description "D = $d_1,\ldots, d_n$", even though many of them do not actively participate in the production of the particular sensation (i.e. do not actively participate in the bringing about of the manifestation of *D*). And there is nothing philosophically or factually questionable about this.

The presented theory of the constitution of mental states is not yet complete, for there is one philosophically problematic issue left to tackle. I argued above that the auditory system has a disposition *D* to produce non-conceptualized auditory sensations. I also claimed that the manifestation of *D* occurs only in the case where a set of simpler dispositions $d_1,\ldots,d_n$ are realized in the programmed way. The manifestation of *D* does not necessarily have to designate the production of only one specific sensation. The description can be applied to all

non-conceptualized auditory sensations. The only difference between the realizations of different sounds is that in each case the set of simpler dispositions $d_1,\ldots, d_n$ contains a peculiar combination of active and latent functions. Nevertheless, if it is claimed that the disposition *D* in question is identifiable with the set of simpler dispositions $d_1,\ldots, d_n$ in the sense that the manifestation of *D* is the programmed realization of $d_1,\ldots, d_n$, at least a few words ought to be said about what is meant by the disposition *D*.

Let's consider the general notion according to which the auditory system *S* has the disposition *D* to produce non-conceptualized auditory sensations from certain kinds of air-pressure changes. According to a functionalist interpretation the situation might be expressed the following way:

the auditory system *S* has a disposition *D* = *S* has some property which plays a particular causal role.

It was already noted that this definition has two major flaws. Firstly, functional characterizations give a false picture of what is going on in the brain during conscious experience: the auditory system does not possess a single physical property at a certain time, which could play any kind of causal role. Secondly, non-conceptualized auditory sensations do not necessarily elicit any kind of behavioral responses or subsequent mental states. Therefore, there is no input-output relation between which the property could function as a causal mediator. But how can disposition *D* then be characterized?

One possibility is to adapt the approach of Stephen Mumford (1998, p.203) and assume a distinction between *concrete* and *abstract* dispositions. Concrete dispositions are the ones we have been so far discussing. For instance, we may say that a brick has a (concrete) multi-track disposition that includes its being hard, breakable etc. However, we may also imagine a situation in which the very same brick is used in the building of the Chinese Wall. In this latter case, the brick can be said to have (of course besides the multi-track disposition) the function of being a part of the Chinese Wall. This new function is not understood in a way that would say that the brick has a disposition to do something. Instead, it is merely a function that is determined by convention. It is relative to minds in the sense that the actual world forms a system, which consists of our general notions, knowledge, used languages, cultures etc., within which the brick can be said to possess the function in question. If no human being were left on earth, there would not be such concepts as ”Chinese Wall” or ”brick”, and the brick would naturally lose its function of being a part of the Chinese Wall. These sort of functions, which are essentially dependent on our responses to certain objects or symbols, can be described as abstract dispositions.

I am inclined to believe that mental states are abstract dispositions; that is, that they exist only relative to minds, which talk about, study, and refer to them. Without such consciousnesses mentality would cease to exist altogether. The idea of mental states as abstract dispositions is very similar to the concept of supervenient causation. Although we constantly say in our ordinary language that the heating caused the boiling of the water, the only causal link that exists is the physical one between the increase of kinetic energy and the bursting of water molecules in the air. The link between heating and boiling is only superveniently causal, which means that it exists only as a figure of speech. We just happen to find it easier to use ordinary language instead of explaining every insignificant phenomena of every-day life in terms of chemistry or physics. However, the fact is that if no human being were left on earth to use language either as a means of communication or as a means of thinking,

only the physical level phenomena would prevail. Correspondingly, if no humans were left, non-conceptualized auditory sensations would still be experienced by members of a variety of species. Of course, their self-consciousness and their capability of using symbolic language are at such low levels that something like the mind-body problem does not exist for them. Therefore, if mental states are understood as abstract dispositions determined by our conventional usage of language and commonsensical experience, nothing is lost but everything is gained.

The fact that non-conceptualized auditory sensations are fully processed preconsciously gives considerable support to my claim. It is a powerful example of the case in which scientists can explain mental phenomena without refering to actual mental terms or expressions of our ordinary language. However, auditory sensations are experienced by persons after all, which means that if mental terms are dropped the actual sensing has to be explained in neuronal terms. This issue could be put a bit differently by stating that mental states are conscious and neural processes on the other hand are not: Thus, how could something unconscious and physical produce something conscious? I am very aware of the importance of this problem, and it should be admitted that the question sets the final and ultimate criterion for the presented theory regarding the constitution of mental states. Therefore, the rest of this chapter is dedicated to solving this problem. I will present some radical and genuinely new ways of handling the issue.

4. The Problem with the Lilliputian Argument: Is Information Processing Conscious?

Ned Block presented in an article entitled ”Troubles with functionalism” (1978) a set of arguments which was primarily directed against Machine Functionalism, but which also – at least indirectly – criticized homunctionalism. Block’s intention was to illustrate that homunculi-headed systems cannot be conscious and possess inner states with qualitative content. At one point Block writes:

> Imagine a body externally like a human body, say yours, but internally quite different. The neurons from sensory organs are connected to a bank of lights in a hollow cavity in the head. A set of buttons connects to the motor-output neurons. Inside the cavity resides a group of little men. Each has a very simple task: to implement a ”square” of a reasonably adequate machine table that describes you...In spite of the low level of intelligence required of each little man, the system as a whole manages to simulate you because the functional organization they have been trained to realize is yours...Through the efforts of the little men, the system realizes the same (reasonably adequate) machine table as you do and is thus functionally equivalent to you. (Block 1978, p.276)

Block’s conclusion is thus that homunctionalist characterizations are mistaken in the sense that they allow one to designate mentality to systems that are obviously not conscious. On the other hand, if the characterization is restricted, homunctionalist descriptions of the human brain are in danger of failing to establish human consciousness.

William G. Lycan (1987a, pp. 28–29; see also 1979;1982) has formulated his own rather technical version of Block’s critique known as ”The New Lilliputtian Argument”. Surprisingly, Lycan’s version ends up with the opposite conclusion to Block’s absent qualia -arguments. The core of Lycan’s argument is that if homunculi are

thought to be intelligent and conscious even at the lowest level, it has the consequence that the homunculi-headed system as a whole is aware of every mental action of the millions of little homunculi that constitute the system. This in turn leads to an unacceptable situation in which a homunculi-headed system possesses thousands of explicitly contradictory beliefs.

Both arguments make valid points, although their factual content can be questioned. For instance, I do not think it justified to accept that if an identical copy of the auditory system were artificially created, it still could not produce auditory sensations. On the other hand, I am not sure how well-founded the claim is that subparts of homunctionalist systems should be conscious and intelligent. However, the importance of these arguments is based on the fact that they force us to reflect on these issues. We have to think only of non-conceptualized auditory sensations, which are found to be fully processed preconsciously. Although the information about these stimuli is extracted unconsciously, the stimuli are experienced as actual sensations only when the information reaches the central executive. It was explained that the central executive is not an entity or some concrete system, but only a set of those processes which are found to bring about conscious experience. Now, Block's and Lycan's examples insist that we ask how the processes of the central executive differ from the preconscious processes? What is it about them that causes the conscious experience of the stimuli?

When the arguments of Block and Lycan, which at first might seem mere philosophical jargon, are contrasted against a concrete frame of reference, their core is revealed. I argued that mental states are nothing more than abstract dispositions determined by convention. However, I added that if no self-conscious mind was left on earth, mental states would cease to exist but sounds would still be heard. Thus, non-conceptualized auditory sensations have the intrinsic qualitative feature of being heard. The absent qualia arguments and the new Lilliputtian argument demonstrate that the key issue is to explain how this phenomenon gets created in the brain. The traditional philosophical answer insists that they cannot be explained in neuronal terms, and that they are either emergent or supervenient with respect to the brain. Nevertheless, it was argued that antireductionism is too problematic a position to be defended, and that emergent and supervenient formulations end up in local, type-identical reductions. Therefore, I am going to apply new neuroscientific facts in support of the thesis that intrinsic, qualitative features of non-conceptualized auditory sensations can be explained in purely physical terms. Furthermore, it will be argued that the auditory information processing-system that creates these phenomena is thoroughly unconscious.

5. Explaining Subjectivity and Qualia

One of the aims of this study has been to come to terms with the mind-body problem and possibly introduce some new insights into the debate. Therefore, much of what I have written has been dictated by the reductionism/antireductionism dispute. The pros and cons of each of the views have been extensively considered from different angles. I participated in this philosophical debate by adopting the view according to which only type-identical characterizations of the mental are acceptable. Though, even the adapted reductionism was found to have some serious theoretical defects and substantial difficulties in coping with the present knowledge of brain

functioning. A dispositional account of the constitution of mental states was presented to correct this situation. The proposed theory was found to be strongly supported by neuroscientific facts. However, in spite of all the presented arguments, antireductionists still have one ace left up their sleeves, which could very well make my efforts seem futile. What I am talking about here is the claim that reductionism should be rejected – no matter how convincing the evidence in support of it might be – because subjectivity and qualia cannot be explained in purely physical terms. This is a genuine issue to which every theory of consciousness – irrespective of the area of research they represent – should be able to give some answer. Neither Jaegwon Kim's type-identical reductionism nor Näätänen's model of auditory processing can do that. Since my own dispositional account was more or less founded on these two models, it has not yet provided a satisfying solution to the problem. Therefore, this issue is taken into consideration below. It is duly noted that the success or failure of the proposed dispositional account of the constitution of metal states is dependent upon the validity of the argued proposition. The outcome will be left for the reader to descide.

There is a vast amount of literature on subjectivity and qualia. Unfortunately, most of it is merely generated by philosophers' needs to earn a living rather than by a genuine desire to learn about the relation between consciousness and the brain. Kripke's (1972) argument against the identity theory is probably one of the saddest examples of this attitude. However, the scope of this study is restricted to non-conceptualized auditory sensations, and, therefore, the task at hand is to explain the origin of their subjective and qualitative aspects. This gives us the chance to learn something genuinely new about qualia. According to my knowledge there have been no other explicit attempts to approach the issue from this perspective.

I hold that the issues of subjectivity and qualia are inter-related in the sense that they are two aspects of the same phenomenon. This can be demonstrated with the help of two famous philosophical examples against the identity theory. The first one was put forward by Thomas Nagel in his distinguished article "What is it like to be a bat?" (1974). Nagel argued that qualia (the qualitative, felt aspect) of any type of sensation is accessible from only a subjective point of view, whereas physical properties have no such limitation. According to the example, a bat's experience has a subjective character. In other words a feeling of what it is like to be a bat, which is accessible only to bats but inaccessible to another species such as humans. The key is that even the most complete understanding of the neurophysiology of bats is not sufficient to obtain this knowledge, but that other organisms can achieve it only by undergoing the very same experiences that bats have. This is due to the fact that Nagel holds that subjectivity is a property of what is known. Because this property is exclusive to conscious experience and not to physical properties, sensations cannot be identified with brain states.

The second example is provided by Frank Jackson (1982;1986), whose views are quite similar with those of Nagel. In an article "What Mary didn't know"(1986) Jackson presented the case of a top neurophysiologist of vision, Mary, who knows everything there is to know about human visual experience but who has been kept all her life in a facility in which it was impossoble for her to see colours. When Mary was released from the captivity, she learns something totally new about the world and other people's experience of it. From the fact that no amount of physical information seems to yield knowledge of the felt aspect of conscious experience, Jackson concludes that it has properties (i.e. qualia) not possessed by physical properties. Therefore, conscious experiences cannot be identified with physical properties and processes.

Nagel's and Jackson's arguments seem at first glance quite similar but there is a significant difference between them. Nagel is mainly talking about a way of knowing in the sense that subjectivity is a property of what is known. In plain words, he is claiming that it is logically and factually impossible for other organisms to know what it is like to be a bat, for they can never possess the subjective character peculiar to bats' experience. Correspondingly, human conscious experience has an irreducibly subjective character unexplainable in physical terms: the only way to achieve a complete knowledge of it is to undergo it personally. That part of Nagel's argument which claims that conscious experience has a qualitative aspect can be accepted, but I am not sure if it can be said to be subjective in any sense. For instance, we can imagine a logically possible situation (which may even be realizable in the near future) in which a person can have access to another person's mental life by means of developed virtual reality or some other sort of neuronal connection. When the experience is shared with another person it cannot be said to be subjective anymore in Nagel's original sense. Therefore, the only way in which Nagel's "subjective" is to be understood is as a method of knowing. For instance, we can say that conscious experience is subjective in the sense that a person's sensations are always coloured by his/her past experiences, cultural factors etc. All this is self-evident. However, in this sense subjectivity has no bearing on our example-case of non-conceptualized auditory sensations, for they do not require any kinds of cultural or other kind of pre-existing frames of mind in order to be heard. Conscious sensations can thus be said to be subjective only in the sense that they are experienced by an entity with some sort of mind.

Jackson's argument, in turn, distinguishes itself from that of Nagel's in that it is not about the way of knowing but rather about the nature of knowing; that is, about qualities or properties of the subject's experience. This is definitly a problem that cannot be circumvented so easily. According to the dispositional account I presented earlier, mental states are abstract dispositions determined and maintained by convention and other minds. If the convention vanishes, mental states as figures of speech disappear too. However, as long as at least one consciousness is left (no matter how primitive) non-conceptualized auditory sensations will be heard. The fact that that person might not be able to understand nor use language does not make any difference: mental states do have the qualitative feature of being experienced as such. Now, the key issue is how this phenomenon can be explained. Nagel and Jackson argue (along the rest of antireductionists) that qualia cannot be explained in neuronal terms, because it is an exclusive property of the mental. If we can question the idea of qualia as some sort of mental property and find an alternative way of explaining the phenomenon in question, there is a good chance that reductionism might turn out to be a valid position after all.

I have found two promising starting points from the vast philosophical literature concerning qualia and subjectivity. The proposers of these approaches have both considered the problem of subjectivity as an epistemological question. Another unifying feature between the two authors is probably that they have actually taken some time to learn about the brain, which is a rarity among philosophers of mind. Paul M. Churchland has suggested in *The Engine of Reason, The Seat of the Soul* (1996, pp. 196–198) that subjectivity is produced by the subject's unique causal link to his own nervous system, brain, and sensations. According to Churchland, the subject is "auto-connected" to his own neural network from which he receives "subjectively" experienced information about his physical state. William G. Lycan makes a similar proposition in his book *Consciousness* (1987a, p.81) and repeats it in *Consciousness and Experience* (1996, p. 68). According to this view, subjectivity

is a product of a certain functional or computational state of the subject's brain, which is unique in respect of the brain state of the observer, although both persons' brain states refer to the same objective state of affairs or extension. For instance, when I am complaining about some pain in my body, the functional or computational state of my brain is different from the one of the doctor who is evaluating the situation next to me. Nevertheless, we are both refering to the same extension, which is the very brain state that causes the pain sensation in me.

Although both suggestions are rather vague, their basic idea is crystal clear: subjectivity and qualia are not properties but products of neural states or rather "self-reflective" neural connections. By "self-reflective" we mean that the brain has processes or states that are directed towards its own processing. Neither Churchland or Lycan has developed the idea any further or provided any concrete and specified examples of how such auto-connections are established in the brain or which parts of the brain might be involved in the procedure. Nevertheless, the idea itself is a useful starting point for further considerations. Our only interest here is to find out whether any such auto-connections can be found from the processing of non-conceptualized auditory sensations. Since we already have detailed knowledge of how such processing takes place and which parts of the brain participate in the production of such sounds, the task should not be overwhelming.

When the auditory system is taken as an example-case, evidence of auto-connections can be found. This has also been noted by Georg Henrik von Wright, who has considered hearing in the light of the mind-body problem in his latest work *In the Shadow of Descartes* (1998). According to his view, the subjectivity characteristic of auditory sensations is caused by a subject's immediate experience of certain processes internal to the auditory system: in some sense, the subject can be said to "hear" his/her own nervous system (1998, p.163). This special relation or way of knowing is accessible only to the subject himself and nobody else. Since two subjects cannot share the same nervous system, von Wright claims that auditory sensations are bound to be private and subjective.

As one of the greatest representatives of analytic philosophy, von Wright is understandably relying more on reasoning than neural sciences. Therefore, his claim appears argumentative rather than strong. Nevertheless, any reader can easily detect the same kind of overtones which were apparent in the views of Lycan and Churchland. Besides, von Wright's suggestion is worth mentioning on the basis that it provides a link between the idea of auto-connection and hearing. Now, it is our job to find out if this link can be made more explicit and concrete.

As was mentioned, non-conceptualized auditory sensations are fully processed preconsciously, which means that all the relevant information about the stimuli is extracted and available before they reach awareness. It was also shown that the neural processing and the neural representations involved in the production of perceptions of sounds can be observed, analysed, and even located with a variety of different brain research methods. Now, we can ask how do the observer's (who is carrying out the measurements) brain states and the test-subject's (who is actually doing all the hearing) brain states differ from each other? After all, the situation is such that the observer has accurate knowledge of the test-subject's neural processes, which are sufficient to produce the sensations, and the subject is actually experiencing those very same processes. Why is it then that only the subject is hearing the sounds? The answer is that the subject has a special relation to these preconscious processes: the subject's brain is auto-connected to the preconscious processes, while the observer's brain is not. The beauty of it all is that in

the case of non-conceptualized auditory sensations we know – even very specifically – what constitutes an auto-connection. The presented model of auditory processing made it very clear that stimuli enter consciousness only when the processed information reaches the central executive. Thus, subjectivity and qualia are produced in the auditory system when preconsciously processed information is brought under the attentive control of the central executive.

A critical mind might naturally wonder what is it that switches the attention of the central executive to focus on specific stimuli. Could it be that the initiator of this switch of focus is itself conscious and that the answer would therefore beg the question? No, it could not. It was conclusively shown, that especially in the condition in which the subject's attention was occupied by a reading task, the attention-triggering device was related to activation of either the MMN or N1 generators. Both of these functions are automatic and unconscious. Therefore, conscious experience (as well as subjective and qualitative aspects of it) has a completely unconscious, neuronal origin. Of course, a sharp-minded reader might yet again ask about the cases in which the central executive switches attention voluntarily. Aren't these situations in which the arousal of conscious experience is explained by the occurence of a process that is itself conscious? The answer is again no. The central executive does nothing by itself: Firstly, if preconscious processes do not provide it with already extracted information, it has nothing to process further. The central executive is nothing but a set of higher-level processes, and if there is nothing to be processed, the central executive has a latent function. Secondly, although we speak of the "conscious, capacity-limited" central executive, it is not conscious by itself. As I argued earlier, *conscious experience is produced when the central executive is auto-connected to preconscious processes*; that is, when it receives information for further processing. If such auto-connection does not exist, neither does conscious experience.

The idea of auto-connection has some further consequences. Since mental states are abstract dispositions, figures of speech, determined and sustained by convention and other minds, the only truly conscious states are auto-connections between the central executive and preconscious processes. Therefore, the only criterion or the "mark" of the mental is the existence of such connections. According to my knowledge, no one has explicated this idea before, and therefore I do not know if the existence of such connections can be measured. However, in the simplified situations (such as the reading condition) even the current means and knowledge can provide radically new means of detecting and defining mental states. In such cases, the elicitation of the MMN corresponds fairly closely with unintended attentional switches and the arousal of conscious (auditory) experience. Though, in some other cases the elicitation of the MMN is not an entirely reliable indicator of awareness. As was mentioned, MMN has been found to occur during sleep (Csepé 1987). Also the consumption of alcohol has been found to suppress mismatch negativity of auditory event-related potentials and attention-switches brought about by the MMN generator processes (see e.g. Jääskeläinen 1995). Nevertheless, the elicitations of MMN and N1 can provide at least some measurable indication of the appearence of mental states. Even with the mentioned limitations and problems, it is a great deal more than all the philosophical jargon will ever amount to.

Conclusions

Now that we have completed our journey through the world of consciousness and the human brain, it is time to summarize what was discovered along the way.

The majority of current philosophers of the mind tend to explain the mind-body relation in terms of supervenience. The popularity of supervenience, it was argued, arouse from the failure of the concept of emergence to provide sufficiently strong psychophysical correlations. Emergence also fails to offer a specific explanation for the relationship between the mental and the physical. The concept of supervenience was originally thought to be apt for the purpose, and it soon replaced emergence in the debate. However, supervenience has never been able to fulfil greater expectations. Basically, supervenience does provide stronger psychophysical correlations, but it is unable to explain the mind-body relation in any way. There have also been serious disagreements regarding what the specific formulation of supervenience.

I sided with Jaegwon Kim in that only the ”strong” version of supervenience can establish a firm relationship between the mental and the physical. I also agreed with Kim that strong supervenience implies such strong nomological necessity that a) it eschews the possibility of mental causation, and that b) it facilitates local, type identical reductions. Since the denial of mental causation commits all forms of nonreductionism to epiphenomenalism, which was found to be an unacceptable view, I decided to favour Kim’s reductionist theory. According to Kim, if a characterization is structurally restricted to include only a certain type of system S (e.g. the human brain), the necessity implied by strong supervenience allows the formulation of local, property-to-property bridge laws of the form ”P __M”:

S _ (P __M).

Kim also claimed that bridge laws (”P _ M”) facilitate local reductions of mental states to pure brain states (”P = M”).

However, when Kim’s view and the widely accepted philosophical notions it implies were compared with the neurophysiology of hearing and the auditory processing of non-conceptualized sounds, it became clear that the basic conceptions of the modern philosophy of the mind were not commensurate with the present knowledge of the structure and functioning of the human brain. Firstly, functional characterizations were found to be quite futile in the case of non-conceptualized sounds, which do not necessarily elicit behavioral responses or subsequent mental states. Secondly, the whole idea that mental states were realized by specifiable and localized physical states at a specific time appeared to be simply wrong. Physical realization of mental states extends over time and might be delayed even for several hundred milliseconds. Nevertheless, they are experienced as occurring at a certain time, which in addition is ”felt” to be the same time at which the stimulus was present. However, probably the most difficult finding was that no single part, process, or state of the human brain can solely physically realize any mental activity. Several independent subsystems of the brain participate in the realization of mental states, and none of them is solely sufficient to produce mental states. This is true even of neuronal populations in the cortical area by which consciousness is generally thought to be created.

A dispositional theory of the constitution of mental states was presented. It was suggested that a mental state could be considered as an abstract disposition D of the whole human brain, which gets manifested when a set of much simpler dispositions $d_1,\ldots, d_n$ are realized in a programmed way. Abstract dispositions were argued to exist only relative to minds in the sense that they are nothing more than figures of speech used in ordinary language. The simpler dispositions $d_1, \ldots, d_n$ in turn were argued to be realized capacities of structurally and functionally specified subsystems of the brain. Multiple realization within the brain was explained by the fact that different subparts can realize the same capacities or even take over the performing of a certain function from other subparts. Therefore, subsystems, which normally participate actively in the production of mental states, can occasionally have *latent functions*.

It was argued that consciousness is nothing more than brain activity. The qualitative aspects of conscious experience (i.e. subjectivity and qualia) were found to be produced by a special causal relationship, *an auto-connection* which the brain has with itself. It was suggested that in the case of the auditory system the auto-connection is most likely formed between the central executive and preconscious processes. It was also suggested that an accurate criterion of mental states could be constructed if the occurrences of such auto-connections could be measured. However, it was discovered that in the case of auditorily evoked, non-conceptualized mental states the activations of the attention-triggering N1 and MMN generators could function as relatively reliable indicators of the appearance of mental states – at least in the reading condition in which the subject's attention is directed away from the auditory stimuli.

This study focused on human consciousness. Therefore, the primary interest of the thesis was directed towards the functioning and the structure of the human brain. It should be emphasized that mental states as abstract dispositions can only be possessed by the human brain. This is because only humans are evolved enough to maintain the cultural contexts and conventions needed for the creation of abstract dispositions. Put plainly, homo sapiens is currently the only species whose members are self-reflective and can reflect on and talk about their mental lives. However, this does not rule out the possibility that less-evolved nervous systems or even organisms of inorganic nature could possess consciousness with qualitative contents. It is highly plausible to think that various organisms and even artificial intelligences might be able to create a qualitative character of consciousness by forming unique kinds of auto-connections between their different structural parts.

References

Aaltonen, O., Tuomainen, J., Laine, M., & Niemi, P. (1993) "Cortical differencies in tonal versus vowel processing as revealed by an ERP component called mismatch negativity (MMN)", *Brain and language*, **44**, pp. 139–152.

Abeles, M. & Goldstein, M. (1972) "Responses of single units in the primary auditory cortex of the cat to tones and to tone pairs", *Brain Research*, **42**, pp. 337–352.

Adrian, E.D. (1964) *The Basis of Sensation: The Action of the Sense Organs*, Hafner Publishing Company, New York.

Aidley, D.J. & Stanfield, P.R. (1996) *Ion Channels: Molecules in Action*, Cambridge University Press, Cambridge.

Alexander, S. (1920) *Space, Time, and Deity, 2 Vols.*, Macmillan, London.

Alho, K. (1995) "Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes", *Ear & Hearing*, **16**, pp. 38–51.

Alho, K., Huotilainen, M., Tiitinen, H., Ilmoniemi, R.J., Knuutila, J., & Näätänen, R. (1993) "Memory-related processing of complex sound patterns in human auditory cortex: a MEG study", *NeuroReport*, **4**, pp. 391–394.

Alho, K., Lavikainen, J., Reinikainen, K., Sams, M., & Näätänen, R. (1990) "Event-related brain potentials in selective listening to frequent and rare stimuli", *Psychophysiology*, **27**, pp. 73–86.

Alho, K., Medved, S.V., Pakhomov, S.V., Roudas, M.S., Tervaniemi, M., Reinikainen, K., Zeffiro, T., & Näätänen, R. (1999) "Selective tuning of the left and right auditory cortices during spatially directed attention", *Brain Research. Cognitive Brain Research*, **7**, pp. 335–341.

Alho, K., Paavilainen, P., Reinikainen, K., Sams, M., & Näätänen, R. (1986) "Separability of different negative components of the event-related potential associated with auditory stimulus processing", *Psychophysiology*, **23**, pp. 613–623.

Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., & Näätänen, R. (1990) "Event-related brain potential of human newborns to pitch change of an acoustic stimulus", *Electroencephalography and clinical Neurophysiology*, **77**, pp. 151–155.

Alho, K., Tervaniemi, M., Huotilainen, M., Lavikainen, J., Tiitinen, H., Ilmoniemi, R.J., Knuutila, J., & Näätänen, R. (1996) "Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses", *Psychophysiology*, **33**, pp. 369–375.

Alho, K., Woods, D.L., Algazi, A., Knight, R.T., & Näätänen, R. (1994) "Lesions of frontal cortex diminish the auditory mismatch negativity", *Electroencephalography and clinical Neurophysiology*, **91**, pp. 353–362.

Alho, K., Woods, D.L., Algazi, A., & Näätänen, R. (1992) "Intermodal Selective Attention. II. Effects of attentional load on processing of auditory and visual stimuli in central space", *Electroencephalography and clinical Neurophysiology*, **82**, pp. 356–368.

Armstrong, D.M. (1961) *Perception and the Physical World*, Routledge & Kegan Paul, London, 1970.

Armstrong, D.M. (1968a) "The headless woman and the defence of materialism", *Analysis*, **29**, pp. 48–49.

Armstrong, D.M. (1968b) *A Materialist Theory of Mind*, Routledge & Kegan Paul, London.

Armstrong, D.M. (1973) *Belief, Truth and Knowledge*, Cambidge University Press, Cambridge.

Armstrong, D.M. (1978) *A Theory of Universals*, Vol.2 of *Universals and Scientific Realism*, Cambridge University Press, Cambridge.

Armstrong, D.M. (1980) *The Nature of Mind and Other Essays*, Cornell University Press, Ithaca (New York).

Armstrong, D.M. (1984) "Self-Profile" in Bogdan (ed.)(1984), pp. 3–51.

Armstrong, D.M. & Malcolm, N. (1984) *Consciousness and Causality*, Basil Blackwell, Oxford.

Aulanko, R., Hari, R., Lounasmaa, O.V., Näätänen, R., & Sams, M. (1993) "Phonetic invariance in the human cortex", *NeuroReport*, **4**, pp. 1356–1358.

Ayer, A.J. (1936/1937) "Verification and experience" reprinted in Ayer (ed.)(1959), pp. 228–243.

Ayer, A.J. (ed.)(1959) *Logical Positivism*, The Free Press, Glencoe (Illinois).

Baars, B.J. (1988) *A Cognitive Theory of Consciousness*, Cambridge University Press, New York.

Baars, B.J. (1994) "A global workspace theory of conscious experience" in Revonsuo & Kamppinen (eds.)(1994), pp. 149–171.

Baddeley, A.D. (1986) *Working Memory*, Clarendon Press, Oxford, 1995.

Balota, D.A. (1983) "Automatic semantic activation and episodic memory encoding", *Journal of Verbal Learning and Verbal Behavior*, **22**, pp. 88–104.

Bauer, R.M. (1984) "Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the guilty knowledge test", *Neuropsychologia 22*, **4**, pp. 457–469.

Bechtel, W. (1994) "Levels of description and explanation in cognitive science", *Minds and Machines*, **4**, pp. 1–25.

Bechtel, W. & Abrahamsen, A. (1991) *Connectionism and the Mind*, Blackwell, Cambridge.

Beckerman, A. (1992a) "Introduction: reductive and nonreductive physicalism" in Beckerman, Flohr, & Kim (eds.)(1992), pp. 11–21.

Beckerman, A. (1992b) ”Supervenience, emergence and reduction” in Beckerman, Flohr, & Kim (eds.)(1992), pp. 94–118.
Beckerman, A., Flohr, H., & Kim, J. (eds.)(1992) *Emergence or reduction?: Essays on the Prospect of Non-reductive materialism*, Walter de Gruyter, Berlin.
Bennett, J (1985) ”Adverb-dropping inferences and the Lemmon criterion” in LePore & McLaughlin (eds.)(1985), pp. 193–206.
Blauert, J. (1983) *Spatial Hearing: The Psychophysics of Sound Localization*, The MIT Press, Cambridge (Massachusetts).
Block, N. (1978) ”Troubles with functionalism” reprinted with author’s revisions in Block (ed.)(1980), pp. 268–305.
Block, N. (1979) ”Philosophy of psychology” in Peter D. Asquith & Henry E. Kyburg, Jr. (eds.) *Current Research in Philosophy of Science*, Philosophy of Science Association, East Lansing, pp. 450–462. Reprinted with author’s revisions and a new title ”Introduction: what is philosophy of psychology?” in Block (ed.)(1980), pp. 1–8.
Block, N. (ed.)(1980) *Readings in Philosophy of Psychology*, Vol.1, Harvard University Press, Cambridge.
Block, N. (ed.)(1981) *Readings in Philosophy of Psychology*, Vol.2, Harvard University Press, Cambridge.
Block, N. (1990) ”Can the mind change the world?” in George Boolos (ed.)(1990) *Meaning and Method: Essays in Honor of Hilary Putnam*, Cambridge University Press, Cambridge.
Boden, M.A. (1990) *The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford.
Bogdan, R.J. (ed.)(1984) *D.M. Armstrong*, D.Reidel Publishing Company, Dordrecht.
Bower, C.H. & Hilgard, E.R. (1981) *Theories of Learning*, 5th ed., Prentice-Hall, Englewood Cliffs (N.J.).
Bradley, M.C. (1963) ”Sensations, brain processes and colours”, *Australasian Journal of Philosophy*, **41**, pp. 372–385.
Brandt, R. & Kim, J. (1967) ”The logic of the identity theory”, *Journal of Philosophy*, **64**, pp. 515–537.
Brentano, F. (1874) *Psychology from an Empirical Standpoint*, trans. Antos C. Rancurello, D.B. Terrell, & Linda L. McAlister, Humanities Press, New York, 1973.
Broad, C.D. (1925) *The Mind and its Place in Nature*, 5th ed., Kegan Paul Ltd., London, 1949.
Broadbent, D.E. (1958) *Perception and Communication*, 2nd ed., Pergamon Press, New York, 1966.
Broadbent, D.E. (1984) ”The maltese cross: a new simplistic model for memory”, *The Behavioral and Brain Sciences*, **7**, pp. 55–94.
Bunge, M. (1977) ”Emergence and the mind”, *Neuroscience*, **2**, pp. 501–509.
Bunge, M. (1980) *The Mind-Body Problem: A Psychological Approach*, Pergamon Press, Oxford.
Bunge, M. (1981) *Scientific Materialism*, D. Reidel, Dordrecht.
Buser, P. & Rougeul-Buser, A. (eds.)(1978) *Cerebral Correlates of Conscious Experience*, North-Holland Publishing Company, Amsterdam.
Böttcher-Gandor, C. & Ullsperger, P. (1992) ”Mismatch negativity in event-related potentials to auditory stimuli as function of varying interstimulus interval”, *Psychophysiology*, **29**, pp. 546–550.
Calvin, W. (1990) *The Cerebral Symphony: Seashore reflections on the Structure of Consciousness*, Bantam, New York.
Campbell, K.K. (1970) *Body and Mind*, Doubleday Anchor Books, New York.
Campbell, J. & Pargetter, R. (1986) ”Goodness and fragility”, *American Philosophical Quarterly*, **23**, pp. 155–165.
Carlen, P.L., Wall, P.D., & Steinbach, R. (1978) ”Phantom limbs and related phenomena in recent traumatic amputations”, *Neurology*, **28**, pp. 211–217.
Carlyon, R.P., Darwin, C.J., & Russell, I.J. (eds.)(1992) *Processing of Complex Sounds by the Auditory System*, Philosophical Transactions series B, The Royal Society, London.
Carnap, R. (1929) *The Logical Structure of the World: Pseudoproblems in Philosophy*, trans. Rolf A. George, University of California Press, Berkley & Los Angeles, 1969.
Carnap, R. (1932) ”The elimination of metaphysics through logical analysis of language”, trans. Arthur Pap, reprinted in Ayer (ed.)(1959), pp. 60–81.
Carnap, R. (1932/1933) ”Psychology in physical language” trans. Georg Schick, reprinted in Ayer (ed.)(1959), pp. 165–198.
Carnap, R. (1934) *The Unity of Science*, trans. M. Black, Thoemmes Press, Chippenham (Wiltshire), 1995.
Carnap, R. (1947) *Meaning and Necessity*, 7th ed., University of Chicago Press, Chicago, 1975.
Celesia, G.G. (1976) ”Organization of auditory cortical areas in man”, *Brain*, **9**, pp. 403–414.
Chappell, V.C. (ed.)(1962) *The Philosophy of Mind*, Prentice-Hall Inc., Englewood Cliffs (New Jersey).
Chisholm, R.M. (1955/1956) ”Sentences about believing”, *Proceedings of the Aristotelian Society*, **56**, pp. 125–148.
Chisholm, R.M. (1957) *Perceiving: A Philosophic Study*, Cornell University Press, Ithaca (New York).
Chisholm, R.M. (1964/1965) ”Believing and intentionality: a reply to Mr. Luce and Mr. Sleigh”, *Philosophy and Phenomenological Research*, **25**, pp. 264–269.
Chisholm, R.M. (1976) *Person and Object*, Allen & Unwin, London.

Chisholm, R.M. (1985) ”The structure of state of affairs” in Vermazen & Hintikka (eds.)(1985), pp. 107–114.
Chomsky, N. (1959) ”A review of B.F. Skinner’s *Verbal Behaviour*” reprinted in Block (ed.)(1980), pp. 48–63.
Churchland, P.M. (1979) *Scientific Realism and the Plasticity of Mind*, Cambridge University Press, Cambridge.
Churchland, P.M. (1981) ”Eliminative materialism and the propositional attitudes” reprinted in Lycan (ed.)(1990), pp. 206–223.
Churchland, P.M. (1984) *Matter and consciousness*, The MIT Press, Cambridge (Massachusetts).
Churchland, P.M. (1996) *The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain*, The MIT Press, Cambridge (Massachusetts).
Churchland, P.S. (1986) *Neurophilosophy: Toward a Unified Science of the Mind–Brain*, The MIT Press, Cambridge (Massachusetts).
Churchland, P.S. & Sejnowski, T.J. (1989) ”Neural representation and neural computation” in Lycan (ed.)(1990), pp. 224–252.
Cowan, N. (1984) ”On short and long auditory stores”, *Psychological Bulletin*, **96**, pp. 341–370.
Cowan, N. (1987) ”Auditory sensory storage in relation to the growth of sensation and and acoustic information extraction”, *Journal of Experimental Psychology: Human Perception and Performance*, **13**, pp. 204–215.
Cowan, N. (1988) ”Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system”, *Psychological Bulletin*, **104**, pp. 163–191.
Cowan, N. (1995) *Attention and Memory: An Integrated Framework*, Oxford University Press, Oxford.
Cowan, N., Winkler, I, Teder, W., & Näätänen, R. (1993) ”Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP)”, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, pp. 909–921.
Crich, F. & Koch, C. (1990) ”Towards a neurobiological theory of consciousness”, *Seminars in the Neurosciences*, **2**, pp. 263–275.
Csepé, V., Karmos, G., & Molar, M. (1987) ”Evoked potential correlates of stimulus deviance during wakefulness and sleep in cat: animal model of mismatch negativity”, *Electroencephalography and clinical Neurophysiology*, **66**, pp. 571–578.
Cummins, R. (1975) ”Functional analysis”, *Journal of Philosophy*, **72**, pp. 741–765.
Cummins, R. (1983) *The Nature of Psychological Explanation*, The MIT Press, Cambridge (Massachusetts).
Dallos, P. (1992) ”The active cochlea”, *The Journal of Neuroscience*, **12**, pp. 4575–4585.
Damasio, A.R. (1989) ”Time-locked multiregional retroactivation: a systems-level proposal for the neural subtrates of recall and recognition”, *Cognition*, **33**, pp. 25–62.
Damasio, A.R. (1990) ”Synchronous activation in multiple cortical regions: a mechanism for recall”, *Seminars in the Neurosciences*, **2**, pp. 287–296.
Davidson, D. (1967a) ”Truth and meaning” reprinted in Davidson (1984), pp. 17–36.
Davidson, D. (1967b) ”Causal relations” reprinted in Davidson (1980b), pp. 149–162.
Davidson, D. (1969) ”The individuation of events” reprinted in Davidson (1980b), pp. 163–180.
Davidson, D. (1970a) ”Mental events” reprinted in Davidson (1980b), pp. 207–225.
Davidson, D. (1970b) ”Events as particulars” reprinted in Davidson (1980b), pp. 181–188.
Davidson, D. (1970c) ”How is weakness of the will possible” reprinted in Davidson (1980b), pp. 21–42.
Davidson, D. (1970d) ”Semantics for natural languages” reprinted in Davidson (1984), pp. 55–64.
Davidson, D. (1971a) ”Eternal *vs*. ephemeral events” reprinted in Davidson (1980b), pp. 189–203.
Davidson, D. (1971b) ”Agency” reprinted in Davidson (1980b), pp. 43–62.
Davidson, D. (1973a) ”The material mind” reprinted in Davidson (1980b), pp. 245–259.
Davidson, D. (1973b) ”Radical interpretation” reprinted in Davidson (1984), pp. 125–140.
Davidson, D. (1973c) ”In defence of convention T” reprinted in Davidson (1984), pp. 65–75.
Davidson, D. (1974a) ”Psychology as philosophy” reprinted in Davidson (1980b), pp. 229–239.
Davidson, D. (1974b) ”On the very idea of the conceptual scheme” reprinted in Davidson (1984), pp. 183–198.
Davidson, D. (1976) ”Hempel on explaining action” reprinted in Davidson (1980b), pp. 261–276.
Davidson, D. (1978) ”Intending” reprinted in Davidson (1980b), pp. 83–102.
Davidson, D. (1980a) ”Comments and replies” in Davidson (1980b), pp. 239–244.
Davidson, D. (1980b) *Essays on Action and Events*, Clarendon Press, Oxford.
Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford.
Davidson, D. (1985) ”Reply to Quine on events” in LePore &McLaughlin (eds.)(1985), pp. 172–176.
Davidson, D. (1993) ”Thinking causes” in Heil & Mele (eds.)(1993), pp. 3–17.
Davies, M. & Stone, T. (eds.)(1995) *Folk Psychology*, Blackwell, Oxford (UK).
Davis, M. (1958) *Computability and Unsolvability*, McGraw-Hill, New York.

Dehaene-Lambertz, G. (1997) ”Electrophysiological correlates of categorical phoneme perception in adults”, *NeuroReport*, **8**, pp. 919–924.
Dennett, D.C. (1975) ”Why the law of effect will not go away” reprinted in Dennett (1978).
Dennett, D.C. (1977) ”Critical notice: *The Language of Thought* by Jerry Fodor”, *Mind*, **86**, pp. 265–280; reprinted with a new title ”A cure for the common code?” in Block (ed.)(1981), pp. 64–77.
Dennett, D.C. (1978) *Brainstorms*, The MIT Press, Cambridge (Massachusetts).
Dennett, D.C. (1987) *The Intentional Stance*, The MIT Press, Cambridge (Massachusetts).
Descartes, R. (1983) *The Principle of Philosophy*, translated with explanatory notes V.R. Miller & R.P. Miller, D.Reidel, Dordrecht.
Descartes, R. (1984) *The Philosophical Writings of Descartes*, Vol.2, translated by Hingham, J., Stouthoff, R., & Murdoch, D., Cambridge University Press, Cambridge.
Deutsch, D. (1988) ”Lateralization and sequential relationships in the octave illusion”, *Journal of the Acoustical Society of America*, **83**, pp. 365–368.
Deutsch, J.A. & Deutsch, D. (1963) ”Attention: some theoretical considerations”, *Psychological Review*, **70**, pp. 80–90.
Diamond, I.T., Jones, E.G., & Powell, T.P.S. (1969) ”The projection of the auditory cortex upon the diencephalon and brain stem in the cat”, *Brain Research*, **15**, pp. 305–340.
Donchin, E., Ritter, W., & McCallum, W.C. (1978) ”Cognitive psychophysiology: The endogeneous components of the ERP” in Callaway, E., Tueting, P., & Koslow, S.H. (eds.)(1978) *Event-related Brain Potentials in Man*, Academic Press, New York, pp. 349–441.
Dretske, F.I. (1977) ”Laws of nature”, *Philosophy of Science*, **44**, pp. 248–268.
Earman, J. (1986) *A Primer on Determinism*, D. Reidel Publishing Company, Dordrecht.
Edelman, G. (1989) *The Remembered Present: A Biological Theory of Consciousness*, Basic Books, New York.
Elliot, L.L. (1970) ”Pitch memory for short tones”, *Perception and Psychophysics*, **8**, pp. 379–348.
Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998) ”Neural mechanisms of involuntary attention to acoustic novelty and change”, *Journal of Cognitive Neuroscience*, **10**, pp. 590–604.
Evans, E.F. (1992) ”Auditory processing of complex sounds: an overview” in Carlyon, Darwin, & Russell (eds.)(1992), pp. 1–12.
Feigl, H. (1958) ”The ´mental´ and the ´physical´”, *Minnesota Studies in the Philosophy of Science*, **2**, pp. 370–497.
Feigl, H. (1971) ”Some crucial issues of mind-body monism”, *Synthese*, **22**, pp. 245–312.
Feldman, F. (1973) ”Kripke’s argument against materialism”, *Philosophical Studies*, **24**, pp. 416–419.
Feldman, F. (1974) ”Kripke on the identity theory”, *Journal of Philosophy*, **71**, pp. 665–767.
Feldman, F. (1980) ”Identity, necessity, and events” in Block (ed.)(1980), pp. 148–155.
Feyerabend, P. (1963) ”Mental events and the brain”, *Journal of Philosophy*, **60**, pp. 295–296.
Field, H.H. (1978) ”Mental representations” reprinted in Block (ed.)(1981), pp. 78–114.
Fodor, J. (1968) ”The appeal to tacit knowledge in psychological explanation”, *Journal of Philosophy*, **65**, pp. 627–640.
Fodor, J. (1974) ”Special sciences, or the disunity of science as a working hypothesis)”, *Synthese*, **28**, pp. 97–115.
Fodor, J. (1975) *The Language of Thought*, Thomas Y. Cromwell Company, New York.
Fodor, J. (1978) ”Propositional attitudes” reprinted in Block (ed.)(1981), pp. 45–63.
Fodor, J. (1983) *The Modularity of Mind*, The MIT Press, Cambridge (Massachusetts).
Fodor, J. (1989) ”Making mind matter more”, *Philosophical Topics*, **17**, pp. 59–80.
Forbes, B.F. & Moskowitz, N. (1974) ”Projections of auditory responsive cortex in the squirrel monkey”, *Brain Research*, **67**, pp. 239–254.
Foster, J. (1991) *The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind*, Routledge, London.
Foster, J. (1994) ”The token-identity thesis” in Warner & Szubka (eds.)(1994), pp. 299–310.
Føllesdal, D. (1985) ”Causation and explanation: a problem in Davidson’s view on action and mind” in LePore & McLaughlin (eds.)(1985), pp. 311–323.
Galaburda, A. & Sanides, F. (1980) ”Cytoarchitectonic organization of the human auditory cortex”, *The Journal of Comparative Neurology*, **190**, pp. 597–610.
Gardner, H. (1987) *The Mind’s New Science: The History of the Cognitive Revolution*, Basic Books, New York.
Garfield, J.L. (ed.)(1990) *Foundations of Cognitive Science*, Paragon House, New York.
Gazzaniga, M.S. (ed.)(1979) *Handbook of Behavioral Neurobiology*, Vol. 2, Plenum Press, New York.
Gazzaniga, M.S. (ed.)(1995) *The Cognitive Neurosciences*, The MIT Press, Cambridge (Massachusetts).
Gazzaniga, M.S. & Blakemore, C. (eds.)(1975) *Handbook of Psychobiology*, Academic Press, New York.
Geach, P. (1957) *Mental Acts: Their Content and Their Objects*, Routledge & Kegan Paul, London.

Giard, M.H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., & Näätänen, R. (1995) "Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis", *Journal of Cognitive Neuroscience*, **7**, pp. 133–143.

Giard, M.H., Perrin, F., & Jaques-Bouchet, P. (1990) "Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study", *Psychophysiology*, **27**, pp. 627–640.

Giard, M.H., Perrin, F., & Pernier, J. (1991) "Scalp topographies dissociate attentional ERP components during auditory information processing", *Acta Otolaryngol.Supp. 491*, pp. 168–175.

Giard, M.H., Perrin, F., Pernier, J., & Bouchet, P. (1990) "Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study", *Psychophysiology*, **27**, pp. 627–640.

Glanzer, M. & Clark, E.O. (1979) "Cerebral mechanisms of information storage: the problem of memory" in Gazzaniga (ed.)(1979), pp. 465–493.

Globus, G.G. (1997) "Explaining consciousness [*existenz*] in quantum terms" in Pylkkänen, Pylkkö, & Hautamäki (eds.)(1997), pp. 100–107.

Goff, G.D., Matsumiya, Y., Allison, T., & Goff, W.R. (1977) "The scalp topography of human somatosensory and auditory evoked potentials", *Electrencephalography and clinical Neurophysiology*, **42**, pp. 57–76.

Goldman, A.I. (1986) "Interpretation Psychologized" in Davies & Stone (eds.)(1995), pp. 74–99.

Gordon, R.M. (1986) "Folk Psychology as Simulation" in Davies & Stone (eds.)(1995), pp. 60–73.

Gower, B. (ed.)(1987) *Logical Positivism in Perspective: Essays on "Language, Truth and Logic"*, Groom Helm, London & Sydney.

Grimes, T.R. (1988) "The myth of supervenience", *Pacific Philosophical Quarterly*, **69**, pp. 152–160.

Hare, R.M. (1952) *The Language of Morals*, Clarendon Press, Oxford.

Hari, R. (1995) "Illusory directional hearing in humans", *Neuroscience Letters*, **189**, pp. 29–30.

Hari, R., Hämäläinen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, J., Salminen, J., Alho, K., Näätänen, R., & Sams, M. (1984) "Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: neuromagnetic recordings in man", *Neuroscience Letters*, **50**, pp. 127–132.

Hari, R. & Loveless, N. (1997) "Hearing backwards in time" in Pylkkänen, Pylkkö, & Hautamäki (eds.)(1997), pp. 168–170.

Harman, G. (1970) "Language learning" reprinted in Block (ed.)(1981), pp. 38–44.

Harré, R. & Madden, E.H. (1975) *Causal Powers: A Theory of Natural Necessity*, Basil Blackwell, Oxford.

Harrison, J.M. (1978) "Functional properties of the auditory system of the brain stem" in Masterton (ed.)(1978), pp. 409–4

Hartman, E. (1977) *Substance, Body and Soul*, Princeton University Press, Princeton.

Hautamäki, A. (1997) "The main paradigms in cognitive science" in Pylkkänen, Pylkkö, & Hautamäki (eds.)(1997), pp. 26–33.

Hawkins, H.L. & Presson, J.C. (1986) "Auditory information processing" in Boff, K.R., Kaufman, L., & Thomas, J.P. (eds.)(1986) *Handbook of Perception and Human Performance*, Wiley, New York, pp. 26.1–26.64.

Hécaen, H. (1979) "Aphasias" in Gazzaniga (ed.)(1979), pp. 239–292.

Heidelberg, H. (1966) "On characterizing the psychological", *Philosophy and Phenomenological Research*, **26**, pp. 529–536.

Heil, J & Mele, A. (1993)(eds.) *Mental Causation*, Oxford University Press, Oxford.

Hellman, G. & Thompson, F. (1975) "Physicalism: ontology, determination, and reduction", *Journal of Philosophy*, **72**, pp. 551–564.

Hempel, C.G. (1935) "The logical analysis of psychology" a revised english version reprinted in Block (ed.)(1980), pp. 14–23.

Hempel, C.G. (1969) "Reduction: ontological and linguistic facets" in Morgenbesser, S. et al. (eds.)(1969) *Philosophy, Science, Method*, New York.

Hempel, C.G. & Oppenheim, P. (1948) "Studies in the logic of explanation" reprinted in Hempel, C.G. (1965) *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, The Free Press, New York, pp. 245–295.

Hiley, B.J. (1997) "Quantum mechanics and the relationship between mind and matter" in Pylkkänen, Pylkkö, & Hautamäki (eds.)(1997), pp. 37–53.

Hill, C.S. (1991) *Sensations: A Defense of Type Materialism*, Cambridge University Press, Cambridge.

Hille, B. (1992) *Ionic Channels of Exitable Membranes*, 2nd ed., Sinauer Associates Inc, Sunderland (Massachusetts).

Hills, D. (1981) "Introduction: mental representations and languages of thought" in Block (ed.)(1981), pp. 11–20.

Hogan, T. & Woodward, J. (1995) "Folk Psychology is here to stay" in Lycan (ed.)(1990), pp. 399–420.

Holender, D. (1986) "Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: a survey and appraisal", *Behavioral and Brain Sciences*, **9**, pp. 1–66.

Honderich, T. (1982) ”An argument for anomalous monism”, *Analysis*, **42**, pp. 59–64.
Horgan, T. (1978) ”The case against events”, *Philosophical Review*, **87**, pp. 28–47.
Horgan, T. (1982) ”Supervenience and microphysics”, *Pacific Philosophical Quarterly*, **63**, pp. 29–43.
Horgan, T & Tye, M. (1985) ”Against the token identity theory” in LePore & McLaughlin (eds.)(1985), pp. 427–443.
Hornsby, J. (1985) ”Physicalism, events, and part-whole relations” in LePore & McLaughlin (eds.)(1985), pp. 444–458.
Hoyle, G. (1975) ”Neural mechanism underlying behavior of invertebrates” in Gazzaniga & Blakemore (eds.)(1975), pp. 3–48.
Hudspeth. A.J. (1989) ”How the ear’s works work”, *Nature*, **341**, pp. 397–404.
Imig, T.J. & Adrián, H.O. (1977) ”Binaural columns in the primary auditory field (A1) of cat auditory cortex”, *Brain Research*, **138**, pp. 241–257.
Jackendoff, R. (1987) *Consciousness and the Computational Mind*, The MIT Press, Cambridge (Massachusetts).
Jackson, F. (1982) ”Epiphenomenal qualia”, *Philosophical Quarterly*, **32**, pp. 127–136.
Jackson, F. (1986) ”What Mary didn’t know”, *The Journal of Philosophy*, **83**, pp. 291–295.
James, W. (1890) *The Principles of Psychology*, Harvard University Press, Cambridge, 1981.
Javitt, D.C., Steinschneider, M., Schroeder, C.E., & Arezzo, J.C. (1996) ”Role of cortical *N*-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: Implications for schizophrenia”, *Proceedings of the National Academy of Sciences of the USA*, **93**, pp. 11962–11967.
Johnston, M. (1985) ”Why having a mind matters” in LePore, & McLaughlin (eds.)(1985), pp. 408–426.
Johnstone, B.M., Patuzzi, R., & Yates, G.K. (1986) ”Basilar membrane measurements and the travelling wave”, *Hearing Research*, **22**, pp. 147–153.
Jouvet, M. (1978) ”Does genetic programming occur during sleep” in Buser & Rougeul-Buser (eds.)(1978), pp. 245–261.
Jääskeläinen, I.P. (1995) *Acute Effect of Ethanol on Attention as Revealed by Event-Related Brain Potentials and Behavioral Measures of Performance*, A Dissertation at the University of Helsinki.
Kahneman, D. & Treisman, A. (1984) ”Changing views of attention and automaticity” in Parasuraman, R. & Davies, R. (eds.)(1984) *Varietes of Attention*, Academic Press, New York, pp. 29–61.
Kallman, H.J. & Massaro, D.W. (1979) ”Similarity effects in backward recognition masking”, *Journal of Experimental Psychology: Human Perception and Performance*, **5**, pp. 110–128-
Kallman, H.J. & Massaro, D.W. (1983) ”Backward masking, the suffix effect, and preperceptual storage”, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, pp. 312–327.
Kandel, E.R., Schwartz, J.H., & Jessell, T.M. (1991) *Principles of Neural Science*, 3rd ed., Elsevier, New York.
Kelly, J.P. (1991) ”Hearing” in Kandel, Schwartz, & Jessell (eds.)(1991), pp. 481–499.
Kemp, D.T. (1978) ”Stimulated acoustic emissions from within the human auditory system”, *Journal of the Acoustical Society of America*, **64**, pp. 1386–1391.
Kessel, R.G. & Kardon, R.H. (1979) *Tissues and Organs: A Text-Atlas of Scanning Electron Microscopy*, W.H. Freeman and Company, San Francisco.
Kim, J. (1971) ”Materialism and the criteria of the mental”, *Synthese*, **22**, pp. 323–345.
Kim, J. (1972) ”Phenomenal properties, psychophysical laws, and the identity theory”, *Monist*, **56**, pp. 177–192.
Kim, J. (1973) ”Causation, nomic subsumption and the concept of event” reprinted in Kim (1993c), pp. 3–21.
Kim, J. (1974) ”Noncausal connections” reprinted in Kim (1993c), pp. 22–32.
Kim, J. (1976) ”Events as property-exemplifications” reprinted in Kim (1993c), pp. 33–52.
Kim, J. (1979) ”Causality, identity and supervenience in the mind-body problem”, *Midwest Studies in Philosophy*, **4**, pp. 31–49.
Kim, J. (1982) ”Psychophysical supervenience” reprinted in Kim (1993c), pp. 175–193.
Kim, J. (1984a) ”Concepts of supervenience”, *Philosophy and Phenomenological Research*, **45**, pp. 153–176.
Kim, J. (1984b) ”Supervenience and supervenient causation”, *The Southern Journal of Philosophy*, **22**, supp., pp. 45–56.
Kim, J. (1984c) ”Epiphenomenal and supervenient causation”, *Midwest Studies in Philosophy*, **9**, pp. 257–270.
Kim, J. (1985) ”Psychophysical Laws” reprinted in Kim (1993c), pp. 194–215.
Kim, J. (1987) ”´Strong´ and ´global´ supervenience revisited”, *Philosophy and Phenomenological Research*, **48**, pp. 315–326.
Kim, J. (1988a) ”Supervenience for multiple domains” reprinted in Kim (1993c), pp. 109–130.
Kim, J (1988b) ”What is ´naturalized´ epistemology?” reprinted in Kim (1993c), pp. 216–236.
Kim, J (1989a) ”Mechanism, purpose, and explanatory exclusion” reprinted in Kim (1993c), pp. 237–264.
Kim, J. (1989b) ”The myth of non-reductive materialism” reprinted in Kim (1993c), pp. 265–284.
Kim, J. (1990) ”Supervenience as a philosophical concept”, *Metaphilosophy*, **21**, pp. 1–27.
Kim, J. (1991) ”Dretske on how reasons explain behaviour” reprinted in Kim (1993b), pp. 285–308.

Kim, J. (1992a) "´Downward causation´ in emergentism and non-reductive physicalism" in Beckerman, Flohr, & Kim (eds.)(1992), pp. 119–138.
Kim, J. (1992b) "Multiple realization and the metaphysics of reduction", *Philosophy and Phenomenological Research*, **52**, pp. 1–26.
Kim, J. (1993a) "Can supervenience and 'non-strict laws' save anomalous monism" in Heil & Mele (eds.)(1993), pp. 19–26.
Kim, J. (1993b) "The nonreductivist's troubles with mental causation" in Heil & Mele (eds.)(1993), pp. 189–210.
Kim, J. (1993c) *Supervenience and Mind: Selected Philosophical Essays*, Cambridge University Press, Cambridge.
Kim, J. (1996) *Philosophy of Mind*, Westview Press, Boulder.
Kimura, D. (1964) "Left-right differencies in the perception of melodies", *Quarterly Journal of Experimental Psychology*, **16**, pp. 355–358.
Kirk, R. (1974) "Zombies *v*. materialists", *Aristotelian Society*, supp., **48**, pp. 135–152.
Knowles, M.C. (1981) "Some remarks on the intentionality of thought", *Philosophy and Phenomenological Research*, **41**, pp. 267–279.
Kosslyn, S.M. (1994) "On cognitive neuroscience", *Journal of Cognitive Neuroscience*, **6**, pp. 297–303.
Kraemer, E.R. (1984) "Divine omniscience and criteria of intentionality", *Philosophy and Phenomenological Research*, **45**, pp. 131–135.
Kraus, N., McGee, T., Carrell, T.D., & Sharma, A. (1997) "Neurophysiologic bases of speech discrimination", *Ear & Hearing*, **16**, pp. 19–37.
Kripke, S. (1971) "Identity and necessity" in M. Munitz (ed.)(1971) *Identity and Individuation*, New York University Press, New York, pp. 135–164.
Kripke, S. (1972) "Naming and necessity" in D. Davidson & G. Harman (eds.)(1972) *Semantics of Natural Language*, D. Reidel, Dordrecht, pp. 253–355.
Landesman, C. (1964) "Mental events", *Philosophy and Phenomenological Research*, **24**, pp. 307–317.
LeDoux, J.E. (1979) "Perietooccipital symptomology: the split-brain perspective" in Gazzaniga (ed.)(1979), pp. 61–74.
Lehtonen, J.B. (1973) "Functional differentiation between late components of visual evoked potentials recorded at occiput and vertex", *Electroencephalography and clinical Neurophysiology*, **35**, pp. 75–82.
LePore E. (1985) "The semantics of action, event, and singular causal sentences" in LePore & McLaughlin (eds.)(1985), pp. 162–171.
LePore, E. & Loewer, B. (1989) "More on making mind matter", *Philosophical Topics*, **17**, pp. 175–191.
LePore, E. & McLaughlin, B. (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, Padstow (Cornwall).
Levin, M.A. (1975) "Kripke's argument against the identity thesis", *The Journal of Philosophy*, **63**, pp. 149–167.
Levinson, A. (1983) "An epistemic criterion of the mental", *Canadian Journal of Philosophy*, **13**, pp. 389–407.
Levänen, S. (1996) *Sensory Memory Traces in the Human Auditory Cortex: Neuromagnetic Studies*, Academic Dissertation at the University of Helsinki.
Levänen, S., Ahonen, A., Hari, R., McEvoy, L., & Sams, M. (1996) "Deviant auditory stimuli activate left and right auditory cortex differently", *Cerebral Cortex*, **6**, pp. 288–296.
Lewes, G.H. (1875) *Problems of Life and Mind, Vol.2.*, Kegan Paul, Trench, Turbner & Co, London.
Lewis, D. (1966) "An argument for the identity theory" reprinted in Lewis (1983), pp. 99–107.
Lewis, D. (1970) "How to define theoretical terms" reprinted in Lewis (1983), pp. 78–95.
Lewis, D. (1972) "Psychophysical and theoretical identifications" reprinted in Block (ed.)(1980), pp. 207–215.
Lewis, D. (1980) "Mad pain and martian pain" reprinted in Lewis (1983), pp. 122–130.
Lewis, D. (1983) *Philosophical Papers*, Vol.1, Oxford University Press, New York.
Lewis, H.A. (1985) "Is the mental supervenient on the physical?" in Vermazen & Hintikka (eds.)(1985), pp. 159–172.
Libet, B. (1978) "Neuronal vs. subjective timing for a conscious sensory experience" in Buser & Rougeul-Buser (eds.)(1978), pp. 131–138.
Loar, B. (1981) *Mind and Meaning*, Cambridge University Press, Cambridge.
Lycan, W.G. (1969) "On intentionality and the psychological", *American Philosophical Quarterly*, **6**, pp. 305–311.
Lycan, W.G. (1974) "Kripke and the materialists", *Journal of Philosophy*, **71**, pp. 677–689.
Lycan, W.G. (1979) "A new lilliputian argument against machine functionalism", *Philosophical Studies*, **35**, pp. 279–287.
Lycan, W.G. (1982) "The moral of the new lilliputtian argument" in Lycan (1987a).
Lycan, W.G. (1987a) *Consciousness*, The MIT Press, Cambridge (Massachusetts).
Lycan, W.G. (1987b) "The continuity of levels of nature" in Lycan (ed.)(1990), pp. 77–96.
Lycan, W.G. (ed.)(1990) *Mind and Cognition: A Reader*, Basil Blackwell, Great Britain.
Lycan, W.G. (1996) *Consciousness and Experience*, The MIT Press, Cambridge (Massachusetts).

Lyytinen, H., Blomberg, A.-P., & Näätänen, R. (1992) "Event-related potentials and autonomic responses to a change in unattended auditory stimuli", *Psychophysiology*, **29**, pp. 523–534.
Lü, Z-L., Williamson, S., & Kaufman, L. (1992a) "Behavioral lifetime of human auditory sensory memory predicted by physiological measures", *Science*, **258**, pp. 1668–1670.
Lü, Z-L., Williamson, S., & Kaufman, L. (1992b) "Human auditory primary and association cortex have different lifetimes for activation traces", *Brain Research*, **572**, pp. 236–241.
Macdonald, C. (1989) *Mind-Body Identity Theories*, Routledge, London.
Mackie, J.L. (1973) *Truth, Probability, and Paradox*, Oxford University Press, Oxford.
Mackie, J.L. (1974) *The Cement of the Universe: A Study of Causation*, Oxford, ?????.
Mackie, J.L. (1977) "Dispositions, grounds, and causes", in Tuomela (ed.)(1978), pp. 99–107.
McDowell, J. (1985) "Functionalism and anomalous monism" in Lepore & McLaughlin (eds.)(1985), pp. 387–398.
McGinn, C. (1977) "Anomalous monism and Kripke's cartesian intuitions", *Analysis*, **37**, pp. 78–80.
McGinn, C. (1978) "Mental states, natural kind, and psychophysical laws", *Proceedings of the Aristotelian Society*, supp. **52**, pp. 195–221.
McGinn, C. (1980) "Functionalism and phenomenalism: a critical note", *Australasian Journal of Philosophy*, **58**, pp. 35–46.
McLaughlin, B.P. (1992) "The rise and fall of British Emergentism" in Beckerman, Flohr, & Kim (eds.)(1992), pp. 49–93.
Margolis, J. (1978) *Persons and Minds: The Prospects of Nonreductive materialism*, D.Reidel, Dordrecht.
Marshall, J.C. (1984) "Multiple perspectives on modularity", *Cognition*, **17**, pp. 209–242.
Massaro, D.W. (1972) "Perceptual images, processing time, and perceptual units in auditory perception", *Psychological Review*, **79**, pp. 124–145.
Massaro, D.W. (1975) *Experimental Psychology and Informtion Processing*, Rand McNally, Chicago.
Masterton, R.B. (ed.)(1978) *Handbook of Behavioral Neurobiology*, vol. 1, Plenum Press, New York.
Matthews, R. (1994) "The measure of mind", *Mind*, **103**, pp. 131–146.
Mellor, D.H. (1974) "In defense of dispositions", *Philosophical Review*, **83**, pp. ???
Mill, J.S. (1843) *A System of Logic* reprinted in *Collected Works, Vol. VII*, Toronto, 1973.
Moore, B.W. (1973) "Brain-specific proteins" in Schneider et al. (1973), pp. 13–26.
Moore, G.E. (1922) *Pricipia Ethica*, Cambridge.
Moray, N. (1959) "Attention in dichotic listening: affective cues and the influence of instructions", *Quarterly Journal of Experimental Psychology*, **11**, pp. 56–60.
Morgan, C.L. (1923) *Emergent Evolution*, William & Norgate, London.
Morton, J., Crowder, R.G., & Prussin, H.A. (1971) "Experiments with the stimulus suffix effect", *Journal of Experimental Psychology*, **9**, pp. 169–190.
Moscovitch, M. (1979) "Information processing and the cerebral hemispheres" in Gazzaniga (ed.)(1979), pp. 379–446.
Mucciolo, L.F. (1973) "Comment: Feyerabend on the identity theory", *Mind*, **82**, pp. 111–112.
Mucciolo, L.F. (1974) "The identity theory and criteria of the mental", *Philosophy and Phenomenological Research*, **35**, pp. 167–180.
Mumford, S. (1998) *Dispositions*, Oxford University Press, Oxford.
Mäkinen, S., Hartikainen, K., Eriksson, J.-T., & Jäntti, V. (1995) "Complex cognition patterns during general anaesthesia" in Pylkkänen & Pylkkö (eds.)(1995), pp. 189–193.
Nagel, E. (1961) *The Structure of Science*, Harcourt, Brace & World, New York.
Nagel, T. (1965) "Physicalism", *Philosophical Review*, **74**, pp. 339–356.
Nagel, T. (1970) "Armstrong on the mind", *Philosophical Review*, **79**, pp. 394–403.
Nagel, T. (1974) "What it is like to be a bat" reprinted in Block (ed.)(1980), pp. 159–170.
Nagel, T. (1986) *The View from Nowhere*, Oxford University Press, New York.
Neisser, U. (1967) *Cognitive Psychology*, Appleton-Century-Crofts, New York.
Neurath, O. (1932/1933) "Protocol sentences" trans. Georg Schick, reprinted in Ayer (ed.)(1959), pp. 199–208.
Nicholls, J.G., Martin, A.R., & Wallace, B.G. (1992) *From Neuron to Brain: A Cellular and Molecular Approach to the Function of the Nervous System*, 3rd ed., Sinauer Associates Inc., Sunderland (Massachusetts).
Niiniluoto, I. (1987) "From possibility to probability: british discussion on modality in the nineteenth century" in Knuuttila, S (ed.)(1987) *Modern Modalities*, Kluwer Academic Publishers, Dordrecht.
Niiniluoto, I. (1994) "Scientific realism and the problem of consciousness" in Revonsuo & Kamppinen (eds.)(1994), pp. 33–54.
Nordby, H., Walton, T., & Pfefferbaum, A. (1988a) "Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals", *Psychophysiology*, **25**, pp. 262–268.
Nordby, H., Walton, T., & Pfefferbaum, A. (1988b) "Event-related potentials to time-deviant and pitch-deviant tones", *Psychophysiology*, **25**, pp. 249–261.

Nottebohm, F. (1979) ”Origins and mechanisms in the establishment of cerebral dominance” in Gazzaniga (ed.)(1979), pp. 295–344.

Novak, G.P., Ritter, W., Vaughn, H.G., & Wiznitzer, M.L. (1990) ”Differentiation of negative event-related potentials in an auditory discrimination task”, *Electroencephalography and clinical Neurophysiology*, **75**, pp. 255–275.

Nozick, R. (1982) *Philosophical Explanations*, Harvard University Press, Cambridge.

Nyman, G., Alho, K., Laurinen, P., Paavilainen, P., Radil, T., Reinikainen, K., Sams, M., & Näätänen, R. (1990) ”Mismatch negativity (MMN) fo sequences of auditory and visual stimuli: evidence for a mechanism specific to the auditory modality”, *Electroencephalography and clinical Neurophysiology*, **77**, pp. 436–444.

Näätänen, R. (1984) ”In search of a short-duration memory trace of a stimulus in the human brain” in Pulkkinen, L. & Lyytinen, P. (eds.)(1984) *Human Action and Personality: Essays in Honour of Martti Takala*, University of Jyväskylä, Jyväskylä, pp. 29–43

Näätänen, R. (1990) ”The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function”, *Behavioral and Brain Sciences*, **13**, pp. 201–288.

Näätänen, R. (1992) *Attention and Brain Function*, N.J. Erlbaum, Hillsdale.

Näätänen, R. (1995) ”The mismatch negativity: a powerful tool for cognitive neuroscience”, *Ear & Hearing*, **16**, pp. 6–18.

Näätänen, R., Gaillard, A.W.K., & Mäntysalo, S. (1978) ”Early selective-attention effect reinterpreted”, *Acta Psychologica*, **42**, pp. 313–329.

Näätänen, R., Jiang, D., Lavikainen, J., Reinikainen, K., & Paavilainen, P. (1993) ”Event-related potentials reveal a memory trace for temporal features”, *NeuroReport*, **5**, pp. 310–312.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997) ”Language-specific phoneme representations revealed by electric and magnetic brain responses”, *Nature*, **350**, pp. 432–434.

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1987) ”The mismatch negativity to intensity changes in an auditory stimulus sequence” in Johnson, R., Rohrbaugh, J.W., & Parasuraman, R. (eds.)(1987) *Current Trends in Event-Related Potential Research (EEG Suppl. 40)*, pp. 125–131.

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1989) ”Do event-related potentials reveal the mechanism of the auditory sensory memory in the brain”, *Neuroscience Letters*, **98**, pp. 217–221.

Näätänen, R., Paavilainen, P., & Reinikainen, K. (1989) ”Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man?”, *Neuroscience Letters*, **107**, pp. 347–352.

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993) ”Attention and mismatch negativity”, *Psychophysiology*, **30**, pp. 436–450.

Näätänen, R. & Picton, T. (1987) ”The N1 wave of the human electric and magnetic response to sound: a review and analysis of the component structure”, *Psychophysiology*, **24**, pp. 375–425.

Näätänen, R., Sams, M., Alho, K., Reinikainen, K., & Sokolov, E.N. (1988) ”Frequency and location specifity of the human vertex N1 wave”, *Electroencephalography and clinical Neurophysiology*, **69**, pp. 523–531.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1995) ”Development of a memory trace for a complex sound in the human brain”, *NeuroReport*, **4**, pp. 503–506.

Oppenheim, P. (1926) *Die Natürliche Ordnung der Wissenschaften: Grundgesetze der vergleichenden Wissenschaftslehre*, Verlag von Gustav Fischer, Jena.

Owens, D. (1989) ”Disjunctive laws”, *Analysis*, **49**, pp. 197–202.

Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1991) ”Right hemisphere dominance of different mismatch negativities”, *Electroencephalography and clinical Neurophysiology*, **78**, pp. 466–479.

Paavilainen, P., Cammann, R., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1987) ”Event-related potentials to pitch change inan auditory stimulus sequence during sleep” in Johnson, R., Rohrbaugh, J.W., & Parasuraman, R. (eds.)(1987) *Current Trends in Event-Related Potential Research (EEG Suppl. 40)*, pp. 246–255.

Paavilainen, P., Jiang, D., Lavikainen, J., & Näätänen, R. (1993) ”Stimulus duration and the sensory memory trace: an event-related potential study”, *Biological Psychology*, **35**, pp. 139–145.

Paavilainen, P., Karlsson, M.-J., Reinikainen, K., & Näätänen, R. (1989) ”Mismatch negativity to change in spatial location of an auditory stimulus”, *Electroencephalography and clinical Neurophysiology*, **73**, pp. 129–141.

Paavilainen, P., Saarinen, J., Tervaniemi, M., & Näätänen, R. (1995) ”Mismatch negativity to changes in abstract sounds features during dichotic listening”, *Journal of Psychophysiology*, **9**, pp. 243–249.

Paavilainen, P., Tiitinen, H., Alho, K., & Näätänen, R. (1993) ”Mismatch negativity to slight pitch changes outside strong attentional focus”, *Biological Psychology*, **37**, pp. 23–41.

Paillard, J., Michel, F., & Stelmach, G. (1983) ”Localization without content: a tactile analogue of blind sight”, *Archives of Neurology*, **40**, pp. 548–551.

Pantev, C., Hoke, M., Lehnertz, K., & Lütkenhöner, B. (1989) "Neuromagnetic evidence of an amplitopic organization of the human auditory cortex", *Electroencephalography and clinical Neurophysiology*, **72**, pp. 225–231.

Penfield, W. & Perot, P. (1963) "The brain's record of auditory and visual experience", *Brain*, **86**, pp. 595–696.

Penrose, R. (1989) *The Emperor's New Mind: Concerning Minds, Computers, and the Laws of Physics*, Oxford University Press, Oxford.

Pereboom, D. & Kornblith, H. (1991) "The metaphysics of irreducibility", *Philosophical Studies*, **63**, pp. 124–145.

Perus, M. (1995) "Analogies between quantum and neural processing: consequences for cognitive science" in Pylkkänen & Pylkkö (eds.)(1995), pp. 115–123.

Peters, A., Palay, S.L., & Webster, H.F. (1970) *The Fine Structure of Nervous System*, Harper & Row, New York.

Pickles, J.O. (1988) *An Introduction to the Physiology of Hearing*, Academic Press, London.

Picton, T.W., Linden, R.D., Hamel, G., & Maru, J.T. (1983) "Aspects of hearing", *Seminars in Hearing*, **4**, pp. 327–341.

Picton, T.W., Stapells, D.R., & Campbell, K.B. (1981) "Auditory evoked potentials from the human cochlea and brainstem", *Journal of Otolaryngology*, **10**, pp. 1–41.

Pihko, E., Leppäsaari, T., Leppänen, P., Richardson, U., & Lyytinen, H. (1997) "Auditory event-related potentials (ERP) reflect temporal changes in speech stimuli", *NeuroReport*, **8**, pp. 911–914.

Pihlström, S. (1996) *Structuring the World: The Issues of Realism and the Nature of Ontological Problems in Classical and Contemporary Pragmatism*, Acta Philosophica Fennica 59, Societas Philosophica Fennica.

Pitson, A.E. (1985) "Frank Jackson and the characterisation of sense-data", *Australasian Journal of Philosophy*, **63**, pp. 428–439.

Place, U.T. (1956) "Is consciousness a brain process?" reprinted in Chappell (ed.)(1962), pp. 101–109.

Pockett, S. (1999) "Anesthesia and the Electrophysiology of Auditory Consciousness", *Consciousness and Cognition*, **8**, pp. 45–61.

Pollock, J.L. (1987) *Contemporary Theories of Knowledge*, Hutchison, London.

Popper, K. (1966) *Of Clouds and Clocks: An Approach to the Problem of Rationality and the Freedom of Man*, Washington University, St.Louis (Missouri),

Popper, K. & Eccles, J.C. (1977) *The Self and its Brain*, Routledge & Kegan Paul, London, 1986.

Posner, M.I., Nissen, M.-J., & Klein, R. (1976) "Visual dominance: an information-processing account of its origins and significance", *Psychological Review*, **83**, pp. 157–171.

Posner, M.I. & Rothbart, M.K. (1990) "Attentional mechanism and conscious experience" in Milner, A.D. & Rugg, M.D. (eds.)(1990) *Neuropsychology of Consciousness*, Academic Press, London, pp. 91–111.

Posner, M.I. & Snyder, C.R.R. (1975) "Attention and cognitive control" in Solso, R.L. (ed.)(1975) *Information Processing and Cognition: The Loyla Symposium*, Lawrence Erlbaum, Hillsdale (N.J.), pp. 55–85.

Post, J. (1987) *The Faces of Existence*, Cornell University Press, Ithaca.

Prior, E. (1985) *Dispositions*, Aberdeen University Press.

Prior, E., Pargetter, R., & Jackson, F. (1982) "Three thesis about dispositions", *American Philosophical Quarterly*, **19**, pp. 251–257.

Pritchard, W.S. (1981) "Psychophysiology of P300", *Psychological Bulletin*, **89**, pp. 506–540.

Putnam, H. (1960) "Minds and machines" in Sydney Hook (1960) *Dimensions of Mind*, New York University Press, New York.

Putnam, H. (1963) "Brains and behaviour" reprinted in Block (ed.)(1980), pp. 24–36.

Putnam, H. (1966) "The mental life of some machines" reprinted in Putnam (1975a).

Putnam, H. (1967) "Psychological predicates" in W.H. Capitan & D.D. Merrill (1967) *Art, Mind, and Religion*, University of Pittsburgh, Pittsburgh; reprinted with a new title "The nature of mental states" in Block (ed.)(1980), pp. 223–231.

Putnam, H. (1975a) *Mind, Language, and Reality: Philosophical Papers*, Vol.2, Cambridge University Press, Cambridge.

Putnam, H. (1975b) "Philosophy of our mental life" reprinted in Block (ed.)(1980), pp. 134–143.

Putnam, H. (1981) *Reason, Truth and History*, Cambridge University Press, Cambridge.

Putnam, H. (1984) "Models and modules", *Cognition*, **17**, pp. 253–264.

Putnam, H. (1988) *Representation and Reality*, The MIT Press, Cambridge (Massachusetts).

Pylkkänen, P. & Pylkkö, P. (eds.)(1995) *New Directions in Cognitive Science*, Finnish Artificial Intelligence Society, Helsinki.

Pylkkänen, P., Pylkkö, P., & Hautamäki, A. (eds.)(1997) *Brain, Mind and Physics*, Frontiers in Artificial Intelligence and Applications –Series, Vol. 33, IOS Press, Amsterdam.

Quine, W.V. (1985) "Events and reification" in LePore & McLaughlin (eds.)(1985), pp. 162–171.

Renault, B., Signoret, J.L., Debruille, B., Breton, F., & Bolgert, F. (1989) "Brain potentials reveal covert facial recognition in prosopagnosia", *Neuropsychologia 27*, **7**, pp. 905–912.

Revonsuo, A., Kamppinen, M., & Sajama, S. (1994) ”General introduction: the riddle of consciousness” in Revonsuo & Kamppinen (eds.)(1994), pp. 1–31.
Revonsuo, A. (1994) ”In search of the science of consciousness” in Revonsuo & Kamppinen (eds.)(1994), pp. 249–285.
Revonsuo, A. (1997) ”Consciousness and levels of description in cognitive science” in Pylkkänen, Pylkkö, & Hautamäki (eds.)(1997), pp. 159–167.
Revonsuo, A. & Kamppinen, M. (eds.)(1994) *Consciousness in Philosophy and Cognitive Neuroscience*, Lawrence Erlbaum Associates, Hillsdale (New Jersey).
Rey, G. (1997) *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Blackwell, Cambridge (Massachusetts).
Rizzo, M., Hurtig, R., & Damasio, A.R. (1987) ”The role of scanpaths in facial recognition and learning”, *Annals of Neurology*, **22**, pp. 41–45.
Romani, G.L., Williamson, S.J., & Kaufman, L. (1982) ”Tonotopic organization of the human auditory cortex”, *Science*, **216**, pp. 1339–1340.
Rorty, R. (1980) *Philosophy and the Mirror of Nature*, 5th ed., Basil Blackwell, Breat Britain, 1989.
Rosen, S. (1992) ”Temporal information in speech: acoustic, auditory and linguistic aspects” in Carlyon, Darwin & Russell (eds.)(1992), pp. 367–373.
Rosenberg. A. (1985) ”Davidson’s unintended attack on psychology” in LePore & McLaughlin (eds.)(1985), pp. 399–407.
Rosenthal, D.M. (1984) ”Armstrong’s causal theory of the mind” in Bogdan (ed.)(1984), pp. 79–120.
Ruggero, M.A., Robles, L., Rich, N.C., & Recio, A. (1992) ”Basilar membrane responses to two-tone stimuli and broadband stimuli” in Carlyon, Darwin, & Russell (eds.)(1992), pp. 307–315.
Russell, B. (1924) ”Logical atomism” reprinted in Ayer (ed.)(1959), pp. 31–50.
Ryle, G. (1949) *The Concept of Mind*, 4th ed., Hutchinson, London, 1951.
Saarinen, J., Paavilainen, P., Schröger, E., Tervaniemi, M., & Näätänen, R. (1992) ”Representations of abstract attributes of auditory stimuli in the human brain”, *NeuroReport*, **3**, pp. 1149–1151.
Sams, M., Hari, R., Rif, J., & Knuuttila, J. (1993) ”The human auditory sensory memory trace persists about 10 s: neuromagnetic evidence”, *Journal of Cognitive Neuroscience*, **5**, pp. 363–370.
Sams, M., Hämäläinen, M., Antervo, A., Kaukoranta, E., Reinikainen, K., & Hari, R. (1985) ”Cerebral neuromagnetic responses evoked by short auditory stimuli”, *Electroencephalography and clinical Neurophysiology*, **61**, pp. 254–266.
Sams, M., Kaukoranta, E., Hämäläinen, M., & Näätänen, R. (1991) ”Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses”, *Psychophysiology*, **28**, pp. 21–29.
Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985) ”Auditory frequency discrimination and event-related potentials”, *Electroencephalography and clinical Neurophysiology*, **62**, pp. 437–448.
Sanford, D.H. (1984) ”Armstrong’s theory of perception” in Bogdan (ed.)(1984), pp. 55–78.
Savin, H. (1980) ”Introduction: behaviorism” in Block (ed.)(1980), pp. 11–13.
Schacter, D.L., McAndrews, M.P., & Moscovitch, M. (1988) ”Acces to consciousness: dissociations between implicit and explicit knowledge in neuropsychological syndromes” in Weiskrantz, L. (ed.)(1988) *Thought without Language*, Oxford University Press, Oxford, pp. 242–278.
Schlick, M. (1932/1933) ”Positivism and realism” trans. David Rynin, reprinted in Ayer (ed.)(1959), pp. 82–107.
Schlick, M. (1934) ”The foundation of knowledge” trans. David Rynin, reprinted in Ayer (ed.)(1959), pp. 209–227.
Schneider, D.J., Angletti, R.H., Bradshaw, R.A., Grasso, A., & Moore, B.W. (eds.)(1973) *Proteins of the Nervous System*, Raven Press, New York.
Schneider, W. & Shiffrin, R.M. (1977) ”Controlled and automatic information processing: detection, search, and attention”, *Psychological Review*, **84**, pp. 1–66.
Schröger, E. (1994) ”Automatic detection of frequency change is invariant over a large intensity range”, *NeuroReport*, **5**, pp. 825–828.
Schröger, E. (1996) ”A neural mechanism for involuntary attention shifts to changes in auditory stimulation”, *Journal of Cognitive Neuroscience*, **8**, pp. 527–539.
Schröger, E., Näätänen, R., & Paavilainen, P. (1992) ”Event-related potentials reveal how non-attended complex sound patterns are represented by the human brain”, *Neuroscience Letters*, **146**, pp. 183–186.
Seager, W. (1991a) ”Disjunctive laws and supervenience”, *Analysis*, **49**, pp. 93–98.
Seager, W. (1991b) *Metaphysics of Consciousness*, Routledge, London.
Searle, J. (1983) *Intentionality*, Cambridge University Press, Cambridge.
Searle, J. (1992) *The Rediscovery of Mind*, The MIT Press, Cambridge (Massachusetts).
Sellars, W. (1956) ”Empiricism and the philosophy of mind” in *Science, Perception and Reality*, Ridgeview Publishing Company, California, 1991.

Shaffer, J.A. (1968) *The Philosophy of Mind*, Prentice-Hall Inc., Englewood Cliffs (N.J.).
Shaffer, J. (1991) ”Mental events and the brain” in D.M. Rosenthal (ed.)(1991) *The Nature of Mind*, Oxford University Press, New York.
Shallice, T. (1984) ”More functionally isolable subsystems but fewer ’modules’”, *Cognition*, **17**, pp. 243–252.
Shelanski, M.L. (1973) ”Microtubules” in Schneider et al. (1973), pp. 227–242.
Skinner, B.F. (1953) *Science and Human Behavior*, Macmillan, New York.
Smart, J.J.C. (1959) ”Sensations and brain processes”, reprinted in Chappell (ed.)(1962), pp. 160–172.
Smart, J.J.C. (1978) ”The content of physicalism”, *Philosophical Quarterly*, **28**, pp. 339–341.
Smart, J.J.C. (1985) ”Davidson’s minimal materialism” in Vermazen & Hintikka (eds.)(1985), pp. 173–182.
Smart, J.J.C. (1994) ”Mind and brain” in Warner & Szubka (eds.)(1994), pp. 19–23.
Smith, A.D. (1977) ”Dispositional properties”, *Mind*, **86**, pp. 439–445.
Smith, B. (1987) ”Austrian origins of logical positivism” in Gower (ed.)(1987), pp. 35–68.
Smith, L.D. (1986) *Behaviorism and Logical Positivism: A Reassessment of the Alliance*, Standford University Press, Stanford (California).
Smolensky, P. (1988) ”On the proper treatment of connectionism”, *Behavioral and Brain Sciences*, **11**.
Sober, E. (1984) *The Nature of Selection*, The MIT Press, Cambridge (Massachusetts).
Sosa, E. (1993) ”Davidson’s thinking causes” in Sosa & Tooley (eds.)(1993), pp. 41–50.
Sosa, E. & Tooley, M. (eds.)(1993) *Causation*, Oxford University Press, Oxford.
Sperling, G. (1960) ”The information available in brief visual presentations”, *Psychological Monographs*, whole no. **498**.
Sperry, R. (1980) ”Mind-brain interaction: mentalism, yes; dualism, no”, *Neuroscience*, **5**, pp. 195–206.
Stephan, A. (1992) ”Emergence: a systematic view on its historical facets” in Beckerman, Flohr, & Kim (eds.)(1992), pp. 25–48.
Stich, S. (1981) ”Dennett on intentional systems”, *Philosophical Topics*, **12**, pp. 39–62.
Stich, S. (1983) *From Folk Psychology to Cognitive Science: The Case Against Belief*, The MIT Press, Cambridge (Massachusetts).
Strawson, P.F. (1985) ”Causation and explanation” in Vermazen & Hintikka (eds.)(1985), pp. 115–135.
Summerfield, Q. & Culling, J.F. (1992) ”Auditory segregation of competing voices: absence of effects of FM or AM coherence”, in Carlyon, Darwin & Russell (eds.)(1992), pp. 63–72.
Szentágothai, J. (1978) ”The local neuronal apparatus of the cerebral cortex” in Buser & Rougeul-Buser (eds.)(1978), pp. 131–138.
Teller, P. (1984) ”Comments on Kim’s paper”, *The Southern Journal of Philosophy*, **22**. *Supplement. Spindel Conference 1983: The Concept of Supervenience in Contemporary Philosophy*, pp. 57–61.
Teller, P. (1992) ”A contemporary look at emergence” in Beckerman, Flohr, & Kim (eds.)(1992), pp. 139–153.
Tervaniemi, M., Alho, K., Paavilainen, P., Sams, M., & Näätänen, R. (1993) ”Absolute pitch and event-related brain potentials”, *Music Perception*, **10**, pp. 305–316.
Thalberg, I. (1985) ”A world without events?” in Vermazen & Hintikka (eds.)(1985), pp. 137–155.
Tiitinen, H., Alho, K., Huotilainen, M., Ilmoniemi, R.J., Simola, J., & Näätänen, R. (1993) ”Tonotopic auditory cortex and the magnetoencephalographic (MEG) equivalent of the mismatch negativity”, *Psychophysiology*, **30**, pp. 537–540.
Trenholme, R. (1978) ”Doing without events”, *Canadian Journal of Philosopy*, **8**, pp. 173–185.
Tuomela, R. (ed.)(1978) *Dispositions*, D.Reidel, Dordrecht.
Tuomela, R. (1994) ”The fate of folk psychology” in Revonsuo & Kamppinen (eds.)(1994), pp. 227–248.
Unger, P. (1976) *Ignorance: A Case for Sceptisism*, Clarendon Press, Oxford.
Valentine, E.R. (1997) ”Deconstructing cognition: towards a framework for exploring non-conceptualized experience” in Pylkkänen, Pylkkö, & Hautamäki (eds.)(1997), pp. 3–12.
Van Gülick, R. (1992) ”Nonreductive materialism and the nature of intertheoretical constraint” in Beckerman, Flohr, & Kim (eds.)(1992), pp. 157–179.
Velmans, M. (1991) ”Is human information processing conscious?”, *The Behavioral and Brain Sciences*, **14**, pp. 651–726.
von Békésy, G. (1960) *Experiments in Hearing*, McGraw Hill, New York.
von Wright, G.H. (1998) *In the Shadow of Descartes: Essays in the Philosophy of Mind*, Kluwer Academic Publishers, Dordrecht.
Warner, R. & Szubka, T. (eds.)(1994) *The Mind-Body Problem: A Guide to the Current Debate*, Basil Blackwell, Oxford.
Warr, W.B. & Guinan, J.J. (1979) ”Efferent innervation of the organ of corti: two separate systems”, *Brain Research*, **173**, pp. 152–155.
Watson, J.B. (1925) *Behaviorism*, 2 nd rewritten and enlarged ed., Kegan Paul, London, 1931.

Waugh, R. (1995) ”Non-conceptualized content: putting meaning before truth” in Pylkkänen & Pylkkö (eds.)(1995), pp. 212–221.

Webster, W.R. & Aitkin, L.M. (1975) ”Central auditory processing” in Gazzaniga & Blakemore (eds.)(1975), pp. 325–364.

Weedman, D., Vause, D., Pongstaporn, T., & Ryugo, D. (1994) ”Corticobulbar synapses in the auditory system: A possible substrate for selective attention”, *Society of Neuroscience Abstracts*, **20**, p.977.

Weiskrantz, L. (1980) ”Varietes of residual experience”, *Quarterly Journal of Experimental Psychology*, **32**, pp. 365–386.

Weiskrantz, L. (1987) ”Residual vision in a scotoma: a follow-up study of ’form’ discrimination” *Brain*, **110**, pp. 77–92.

Weiss, A.P. (1925) *A Theoretical Basis of Human Behavior*, R.G. Adams & Co, Columbus (Ohio).

White, S. (1986) ”Curse of the qualia”, *Synthese*, **68**, pp. 333–368.

Wilson, N. (1974) ”Facts, events, and their identity conditions”, *Philosophical Studies*, **25**, pp. 303–321.

Winter, O., Kok, A., Kenemans, J.L., & Elton, M. (1995) ”Auditory event-related potentials to deviant stimuli during drowsiness and stage 2 sleep”, *Electroencephalography and clinical Neurophysiology*, **96**, pp. 398–412.

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, Kegan Paul, London.

Wittgenstein. L. (1953) *Philosophical Investigations*, trans. G.E.M. Anscombe, Blackwell, Oxford.

Yarvin, H. (1978) ”Criteria of the physical”, *Metaphilosophy*, **9**, pp. 122–132.

Young, A.W. (1994) ”Neuropsychology of awareness” in Revonsuo & Kamppinen (eds.)(1994), pp. 173–203.

Young, A.W. & De Haan, E.H.F. (1990) ”Impairments of visual awareness”, *Mind & Language*, **5**, pp. 29–48.

Yong, E.D., Spirou, G.A., Rice, J.J., & Voigt, H.F. (1992) ”Neural organization and responses to complex stimuli in the dorsal cochlear nucleus” in Carlyon, Darwin, & Russell (eds.)(1992), pp. 407–413.

Zeidel, E. (1978) ”Concepts of cerebral dominance in the split brain” in Buser & Rougeul-Buser (eds.)(1978), pp. 263–284.

Zwislocki, J.J. (1960) ”Theory of temporal auditory summation”, *Journal of the Acoustical Society of America*, **32**, pp. 1046–1060.

Zwislocki, J.J. (1969) ”Temporal summation of loudness: an analysis”, *Journal of the Acoustical Society of America*, **46**, pp. 431–440.