

The Logic and Pragmatics of the Representation and Alteration of Beliefs

Armin Walter Schulz

London School of Economics and Political Science

MPhil Thesis

UMI Number: U615642

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U615642

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

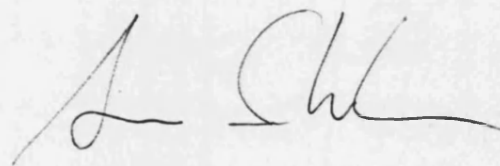
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

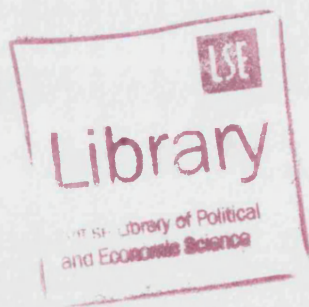
I declare that the work presented in this thesis is my own.

A handwritten signature in black ink, consisting of a stylized 'A' followed by 'Schulz'.

Armin Schulz

THESES

F
8597



1095014

Abstract

In this thesis, I show the extent to which the distinction between logical rationality (the consistency with a set of assumptions) and pragmatic rationality (the strong tendency of providing benefits to actual agents) helps to make sense of probabilistic accounts of the representation and alteration of beliefs.

In order to do this, I first show how the probabilistic representations of beliefs can be seen to follow on from the failure of the cogency of the Logical Theories of probability. I then move on to discuss the four classic theories of probabilistic representations of belief in the literature (those of Ramsey, de Finetti, Savage and Jeffrey) and a key modern treatment (that of Howson & Urbach). Thirdly, I continue the argument by discussing the two key justifications for the core account of the probabilistic alteration of beliefs – Bayesian Conditionalisation – to show that these arguments – if anything – only show the logical rationality of this way of altering beliefs, but not its pragmatic rationality. In a fourth step, I provide a novel justification of this sort by basing it on the tendency of Bayesian Conditionalisation to structure an agent's thoughts and decisions in a way that lowers her decision-making costs. I also discuss some of the consequences of such a justification for Bayesian Conditionalisation, in particular with a view to other conditionalisation principles like Jeffrey Conditionalisation. Finally, I point out some connections of this discussion to contemporary and traditional philosophy of science.

Table of Contents

I. Logic and Pragmatics of the Representation and Alteration of Beliefs	6
1. BELIEFS, THEIR REVISION, AND TWO SENSES OF RATIONALITY	6
1.1. <i>The Two Senses of Rationality</i>	7
1.2. <i>Representations of Beliefs</i>	13
1.3. <i>Alterations of Beliefs</i>	14
1.4. <i>Science and the Probabilistic Representation & Alteration of Beliefs</i>	16
2. TOWARDS A COMPELLING THEORY OF THE REPRESENTATION AND ALTERATION OF BELIEFS: THE PLAN OF THE THESIS.....	17
3. SUMMARY.....	20
II. Historical and Conceptual Origins of the Subjective Theory of Probability: The Logical Theories of Keynes and Carnap.....	22
1. THE LOGICAL THEORY AND THE PRINCIPLE OF INDIFFERENCE	23
2. PROBLEMS AND PARADOXES OF THE PRINCIPLE OF INDIFFERENCE	26
3. SOLUTIONS TO THE PARADOXES AND THEIR SUCCESS.....	31
3.1. <i>Keynes' Solution</i>	32
3.2. <i>Jaynes' Solution</i>	35
4. FROM THE LOGICAL TO THE SUBJECTIVE THEORIES.....	38
III. The Probabilistic Representation of Beliefs: An Overview	42
1. RAMSEY'S "TRUTH AND PROBABILITY"	43
1.1. <i>Ramsey's Method</i>	43
1.2. <i>Criticisms and Replies</i>	50
1.3. <i>Conclusion</i>	57
2. DE FINETTI'S "FORESIGHT: ITS LOGICAL LAWS, ITS SUBJECTIVE SOURCES".....	58
2.1. <i>Dutch Books and the De Finetti Theorem</i>	58
2.2. <i>The Limits of the Betting Approach</i>	62
2.3. <i>Conclusion</i>	73
3. SAVAGE'S "THE FOUNDATIONS OF STATISTICS".....	74
3.1. <i>Savage's Decision Theory</i>	74
3.2. <i>Problems and Criticisms</i>	81
3.3. <i>Conclusion</i>	88
4. JEFFREY'S "THE LOGIC OF DECISION (2ND EDITION)"	90
4.1. <i>Jeffrey's "Logic of Decision"</i>	90
4.2. <i>Criticism and Discussion</i>	97
4.3. <i>Conclusion</i>	107
5. HOWSON & URBACH'S "SCIENTIFIC REASONING – THE BAYESIAN APPROACH" .	108
5.1. <i>Howson & Urbach on the Representation of Beliefs</i>	108
5.2. <i>Problems and Replies</i>	115
5.3. <i>Conclusion</i>	121
6. THE PROBABILISTIC REPRESENTATION OF BELIEFS	123
6.1. <i>The Structures of the Theories</i>	123
6.2. <i>The Rationality of Probabilistic Representations of Belief</i>	125
6.3. <i>Summary and Outlook</i>	130

IV. Bayesian Conditionalisation and the Alteration of Beliefs: An Overview	132
1. THE SOUNDNESS OF BAYESIAN CONDITIONALISATION	133
2. BAYESIAN CONDITIONALISATION: JUSTIFICATIONS AND OBJECTIONS	136
2.1. <i>The Cogency and Centrality of Rule BC: An Argument from Logical Rationality</i>	
.....	137
2.2. <i>The Dutch Book Argument Revisited: A Justification for Bayesian</i>	
<i>Conditionalisation based on Pragmatic Rationality.....</i>	147
3. CONCLUSION.....	153
V. Coherence in Thinking: A Pragmatic Justification for Bayesian	
Conditionalisation	155
1. COHERENTLY STRUCTURED THOUGHTS: THE PRAGMATIC RATIONALITY OF BAYESIAN	
CONDITIONALISATION	156
1.1. <i>A Coherent System of Beliefs.....</i>	158
1.2. <i>The Pragmatic Rationality of a Coherent System of Beliefs.....</i>	163
2. OBJECTIONS AND REPLIES	171
2.1. <i>Two Objections and Two Replies.....</i>	171
2.2. <i>The Failure of Rigidity and Certainty: A Case Study.....</i>	177
3. CONCLUSION.....	186
VI. The Probabilistic Representation and Alteration of Beliefs: Summary and	
Outlook.....	188
1. THE PROBABILISTIC REPRESENTATION & ALTERATION OF BELIEFS AND THE	
PHILOSOPHY OF SCIENCE	188
2. FINAL SUMMARY AND OUTLOOK.....	192
Bibliography.....	194

List of Figures

Figure 1 (Geometrical Paradoxes of the Principle of Indifference)	p. 29
Figure 2 (Geometrical Paradoxes of the Principle of Indifference)	p. 30
Figure 3 (Geometrical Paradoxes of the Principle of Indifference)	p. 30

I. Logic and Pragmatics of the Representation and Alteration of Beliefs

1. Beliefs, Their Revision, and Two Senses of Rationality

In an environment like ours, where being *very* wrong is much worse than being only *slightly* wrong, uncertainty can be a virtue (see also Joyce 1999, p. 579). Given the fact that the badness of error and the goodness of success come in *degrees*, a rational agent will respond, firstly, by having beliefs of various *strengths*, and secondly, by looking for appropriately rational ways of *altering* these beliefs.¹

All of these claims should be immediately obvious: it quite simply *seems* to be the right policy to fit one's cognitive (doxastic) apparatus to the world in this way, and to keep it up-to-date with the evolution of one's interactions with this world. However, whether this is conceded to be obvious or not, I shall not argue for this view of the nature of beliefs here (barring some brief remarks later on in this chapter), but simply take it as a starting *assumption*. Those who are inclined to disagree with its truth should therefore see what follows as an exploration of how much sense can be made of an agent's cognitive attitudes in a framework that is built on shaky foundations (which can still be an interesting investigation). Those who agree with the assumption on the other hand should

¹ In what follows, I use "belief" and "degree of belief" largely interchangeably. This implies that two strengths of the belief that it will rain tomorrow are as different from each other as from the belief that Gordon Brown will be the next Prime Minister. It is possible to draw a more principled distinction, and see "beliefs" as ranging over different *types* of beliefs (i.e. over beliefs that concern different propositions or have different intentional content) and "degrees of belief" as ranging over different strengths of the same belief (i.e. over different degrees of a belief having some particular content). Making this finer distinction however is not relevant in the present context: all the conclusions drawn here could be reformulated with it in mind, which means that my conflation is costless. It does however make for an easier exposition of the points to be discussed.

see the following remarks as drawing out some of its implications and unearthing some of its foundations.

1.1. The Two Senses of Rationality

The notion of “rationality” plays a crucial part in this picture of the importance of degrees of belief and their alteration: it is “rational” agents who have scaled beliefs and alter them by means of “rational” rules. “Rationality” is, of course, a vague and ambiguous term, so that one of the key aims of a theory of beliefs and their alteration has to be sought in making precise exactly what it might mean.

That is, a compelling suchlike theory has to firstly specify the relation between a rational agent and beliefs of various strengths. This in turn implies finding a way of *representing* these beliefs in the theory with sufficient accuracy. Secondly, given this representation, such a theory has to show which rules² of the *alteration* of these beliefs are “rational”, and what exactly this means.³

In order to achieve this, I want to distinguish two senses of "rationality" here: a logical one and a pragmatic one. This distinction is of great relevance in the present thesis and will be one of the key driving forces behind its conclusions, so that it is worthwhile to spend some time here to make it as clear and precise as necessary for a successful start of the inquiry. The full extent of the distinction will however only become obvious over the entire course of the discussion – as with many other distinctions, it is best understood with reference to its *applications*. Since it is the main aim of this work to bring out the senses in which the representation and alteration of beliefs can be said to be “rational”,

² In fact, this theory will also have to justify why these alterations of beliefs require *rules* at all.

³ This is another reason why a representation of these beliefs is necessary: without it, the *formulation* of rules of altering them would not be possible in the first place.

the remarks given here should thus be seen as introductory rather than final. That said, the following sketches of a definition of the two kinds of rationality at the centre of this thesis can be given:

Logical Rationality: Consistency of an ideal reasoner's attitudes and actions with the demands of a predefined system of rules and requirements.

Pragmatic Rationality: The strong tendency of an actual reasoner's attitudes and actions to directly or indirectly yield beneficial consequences.

Let me make a few remarks here so as to throw light on these somewhat opaque defining sketches. Start with the former kind of rationality: the two elements of this definition that need to be further spelled out are, firstly, the use of the term "consistency" and, secondly, the reference to an "ideal reasoner".

"Consistency" here should be understood as it is defined in standard logic.⁴ This can best be illustrated by means of rules for altering beliefs: given a set of formal constraints specified by the theory in question, rules for revising beliefs are logically rational if their results are "sound" i.e. preserve those properties of the beliefs that the logic singles out as significant.⁵ Examples of these properties (which normally will have to be given

⁴ Note that there is some ambiguity in the term "logic": it can either be taken to refer to the entire subject that studies the syntactic and semantic relations between linguistic entities, or as *one particular system within* this subject (e.g. a deduction system that employs only one rule of inference and one that employs two would be seen as – partially – defining different "logics"). In the present thesis, I use the term in its first sense throughout.

⁵ Here, the parallels to deductive logic are clearly visible: for example, a derivation system is "sound" if and only if its axioms together with its rules of inference yield outcomes that are "truth-preserving", that is, outcomes that respect the semantic foundation of deductive logic (nowadays normally spelled out in a Tarskian truth definition); for more on this, see Enderton 2001, p. 131.

exogenously when setting out the theory in question) are the transitivity of preferences entailed by the beliefs in question, and the satisfaction of the laws of the probability calculus.⁶

Secondly, the reference to an “ideal reasoner” implies that there is no requirement that a logically rational updating rule also be practically applicable in some straightforward way. It is rather the case that the rule could *in principle* be used to reach consistent conclusions, which is not to say that it can *actually* be so used. This a situation that is once again parallel to that in standard (propositional) logic: truth tables represent an effective procedure for checking the satisfiability of a set of sentences – which however does not entail that they are also practically applicable in any but the most simple cases (see also Enderton 2001, p. 61-62). “Ideal reasoners” then are those (fictional) agents whose unlimited computing power and memory allows them to reason unconstrained by any physical or biological limitations. They are bound solely by the laws of logic and the exogenous requirements of the system of rules in question.

Moving on to pragmatic rationality (using once more rules of belief revision for purposes of illustration): a pragmatically rational updating rule is a useful heuristic device or a helpful rule of thumb for an actual agent. That is, pragmatically, an updating rule is “rational” if following it *normally* leads to outcomes that are considered appropriate from the point of view of a real human agent. It only needs to do so “normally”, as it is but a fallible heuristic device – which of course does not mean it is unhelpful in practice.

The above definition uses the term “tendency” to describe this property more closely: as argued for Cartwright 1999 (and going back to the work of Mill 1844), a tendency in this sense has to be thought of as an effective *causal* force, but not necessarily as the sole

⁶ For more on this latter kind of requirement, see below.

one. Only in certain circumstances can it be expected to fully determine the outcome of the process in question. However, if this causal factor is strong enough, it will make itself felt in most relevant (for the agent) situations. This last part is also the reason why following such a rule is beneficial to the agent: the strong *tendency* to yield helpful consequences can in many of the situations she will face be counted on to *actually* yield these helpful consequences (alas, it need not to do so in *all* situations however).

The reference to an *actual* agent is important here, as it signals the presence of constraints not relevant for ideal agents. Actual agents are bound by computational limitations, as well as by other physical and biological constraints. They can make mistakes in their ways of reasoning or fail to come to any conclusion whatsoever about the outcome of some deliberation. As these constraints are varied for different actual agents (for example, some have more time to spend deliberating about what to do in certain circumstances than others), this also means that pragmatic rationality is agent-relative. What is pragmatically rational for one person might not be so for another.

Two further remarks are necessary here: firstly, the notion of being “beneficial” in the definition of pragmatic rationality has to be understood relative to an agent-external perspective. That is, an action or attitude is pragmatically rational if it benefits that agent, *whether that agent knows this or not*. One such case is altruistic behaviour: if cooperation with other agents is beneficial (as will be assumed in chapter V for example), then acting altruistically might be pragmatically rational, despite the fact that it might go against the agent’s *internal* best interest (e.g. her immediate desire-satisfaction).

Secondly, despite the fact that pragmatic rationality seems to be ‘reduced to’ little more than beneficiality here, this does not mean that it is not ‘true’ rationality anymore.

This can be brought out most easily by considering the following objection: at first pass, it might seem mistaken to construe something that merely tends to be *beneficial* as *rational* (in any sense). For example, some agent might have a friendly and open nature towards others ever since birth (it might be genetically determined, say), and this attitude moreover *happens* to tend to be beneficial for her. Here, it seems strange to call this attitude ‘rational’ (even only in the pragmatic sense) just because it incidentally also tends to cause beneficial consequences.⁷

However, there are three points to be kept in mind here. Firstly, whether or not we would call such an attitude “rational” in ordinary language is an open question. It might not strain commonsense too much to say that – in some sense of this notoriously vague term – a jovial attitude can be “rational”. Secondly, this will become even clearer if it is noted that the attribution of pragmatic rationality to an *attitude* can be seen as a *hypothetical* claim about the pragmatic rationality of the *adoption* of that attitude. The idea is that if an agent *were to* deliberate about whether to *adopt* a certain kind of attitude or not, she would be acting pragmatically rational in adopting those attitudes that tend to have beneficial consequences. A friendly turn of mind thus can be pragmatically rational because its *adoption* (an action) might be pragmatically rational. For agents who simply *have* that attitude, therefore, the attribution of pragmatic rationality comes about in a roundabout way, through the pragmatic rationality of the hypothetical adoption of the attitude.

Still, this might create a class of rational attitudes or actions that is larger than what we commonsensically might take it to be. However, even if this were the case, it would not be harmful in the present context, where the emphasis is squarely on *one* kind of attitude

⁷ I thank Dan Hausman for this example.

only: beliefs. Thus, even if the above definition included certain attitudes within the class of rational ones commonsense did not, that would not bear on the issues under discussion here.

The third point to keep in mind is that rationality is not *merely* restricted to beneficiality here, but that causal connections are given a large role to play, too. That is, the attitude or action in question must be *causally responsible* – in the sense of a *causal tendency* – for the beneficial consequences to qualify as pragmatically rational. Mere *correlation* between an agent's actions or attitudes and the beneficial consequences is not enough for that: the former would not be connected to the latter closely enough. *Rationality* requires a tight link between an action or attitude and its consequences, and this link is provided here by the former being a causal tendency of the latter. In this way, this sense of rationality is again brought closer to ordinary ways of speaking.

Having thus gained an initial insight into these two different senses of rationality, it is important to be aware of the fact that they are logically independent from each other. For example, some ways of altering beliefs might not be rational in the logical sense, but only in the pragmatic one. This might mean that they are inconsistent or only applicable in certain cases, but that they nonetheless tend to yield successful outcomes – for example, the inconsistency might only make itself felt in situations that are far removed from any of those the particular actual agent will face. On the other hand, following a rule can be pragmatically irrational (e.g. if applying it is impossible for human agents due to constraints of time and place), whilst at the same time being logically rational: in

principle, following it yields outcomes consistent with the desiderata of a particular theory of beliefs, so that it *is* appropriate for an *ideal* agent – but only for him.⁸

1.2. Representations of Beliefs

Given this initial overview of the key distinction of this thesis, it is now possible to start considering the first half of its field of application: the representation of beliefs. Here, the aim is to use the distinction between logical and pragmatic rationality to come to a clear and coherent understanding of the various ways of theoretically representing beliefs. As a matter of fact, there is a whole host of accounts that purport to represent beliefs in the literature; however, they can be distinguished easily as to whether they are probabilistic or non-probabilistic in nature.

The former sees probabilities as representing degrees of belief, whilst the latter tends not to even concede that there *are* degrees of belief (see for example Jeffrey 1985, p. 45). Of course, these two ways of representing beliefs face very different challenges: the former has to justify why degrees of belief should be seen as being representable by probabilities (this task will be surveyed in chapter III). The latter, on the other hand, has to make sense of the appeal of the assumption (stated at the beginning of the thesis) that the value of the consequences of one's interactions with the world comes in degrees, leading to a similarly gradualist cognitive makeup of the agent.

As made clear earlier, the present inquiry is limited entirely to probabilistic representations of belief. This is not necessarily to say that there are no convincing other

⁸ Another example where the two notions of rationality go apart is the following: assume some slightly mischievous philosophy lecturer *rewards* students who *genuinely* accept a vast array of logical fallacies. Then it might be *pragmatically* rational for the students to be *logically* irrational in this way, at least in the classroom. I thank Dan Hausman for this example.

accounts of representing beliefs, but merely to restrict the scope of the thesis: at the very least, probabilistic representations seem to present a promising theoretical framework within which to ask the questions that form the core of the present inquiry. In this way, the present thesis can be read weakly as merely trying to point out a number of issues that *probabilistic* representations (and alterations – see below for more on this) of beliefs have to face, without necessarily committing itself to the stance that this is the only valid way of doing so.⁹ A stronger reading of the thesis would see it as actually *advocating* the probabilistic representation of beliefs. However, with these remarks I shall leave this issue aside and continue using the assumption that beliefs come in degrees, and that these degrees are best represented probabilistically.¹⁰

1.3. Alterations of Beliefs

It seems obvious that, given a probabilistic representation of beliefs, any account of their alteration has to be equally probabilistic. Since the present thesis focuses solely on these kinds of representations of beliefs, it should therefore not come as a surprise that it is also only concerned with probabilistic accounts of their alteration.¹¹

However, I shall restrict myself even further in this case. The present inquiry will focus on only *one* rule for the revision of beliefs – Bayesian Conditionalisation. The

⁹ An important overview over the question of how to best represent beliefs is Jeffrey 1985; unsurprisingly, this work concludes very negatively for non-probabilistic representations of belief. Another take on this matter is Joyce 1998.

¹⁰ Why exactly *probabilities* can be used to represent beliefs forms the core of chapter III.

¹¹ However, once again, it is important to be aware of the fact that just as there are non-probabilistic ways of *representing* beliefs, there are also non-probabilistic ways of *revising* them. An important and notable one amongst these is the Cartesian method of doubt: if some belief turns out to be not “certain and indubitable”, it is to be *dropped* from one’s system. Once again though, and for the same reasons cited above concerning the representation of beliefs, I shall limit myself here to probabilistic rules. Moreover, just as it did above, this need not mean that available alternatives are not viable, but merely that the present thesis is not concern with them. See also Jeffrey 1985, p. 55.

reasons for this are discussed in some detail in chapter IV, but for now, suffice it to say that of the many rules that have been proposed, this one is often taken to be of the greatest importance. Moreover (and as will be made clear below), the rule can also claim to be very intuitive and fruitful. On top of all that, it is probably the most widely known suchlike rule. All of this points to the pivotal position Bayesian Conditionalisation must hold in an inquiry of the present form.¹²

The aspect of Bayesian Conditionalisation that is the main focus of this thesis is the extent to which it can be seen to have a well-grounded justification. Again, this question can be more fruitfully stated by falling back on the above distinction between two kinds of rationality: Is it *rational* to follow the rule? If so, in what sense of “rational”? Setting up the inquiry with this distinction in mind presents many familiar problems in a new light.

One preliminary aspect concerning the alteration of beliefs that has to be mentioned at this point is the following. In general, the origin of such an alteration can be either metaphysical (in the sense of an actual change in the world) or epistemological (as to what we know about this world) in nature. However, it turns out that in what follows this distinction is of little importance: since the focus of the inquiry is on epistemological issues, actual changes in the world can be seen as secondary. The primary focus is on our *learning* something new about the world – independently of whether that is a changed or an unchanged world. This can be illustrated with the following (stylised) example.

Assume that at time t_0 , there is evidence that the world is of type A and at time t_1 , new evidence that it is of type B. Assume further that this new evidence requires the revision

¹² That said, however: I will make brief reference to alternative rules of conditionalisation (especially Jeffrey Conditionalisation) where this is deemed important and helpful for the general thrust of the inquiry.

of beliefs. Then from the point of view of this inquiry, it is irrelevant whether the presence of this new evidence is due to there being a change in the fundamental fabric of the world (so that the world has changed from being of type A to being of type B), or merely due to our having found out something we did not know before (so that we were misled to think of it as of type A at t_0). The present thesis is merely concerned with analysing the *effects* of evidence on beliefs, and not with the *origins* of that evidence.

1.4. Science and the Probabilistic Representation & Alteration of Beliefs

The final aspect concerning the importance of a theory of the rationality of belief representation and revision worth pointing out is its connection with the philosophy of science. Traditionally, much of this debate has been set in the framework of a 'Bayesian' philosophy of science:¹³ if science is seen as an epistemic venture (which seems very plausible), then a methodology for how beliefs about the world are to be represented and altered bears a close relation to an account of the methodology of science.

It should however be noted that the inquiry here is more general: it is concerned with beliefs in general, and its conclusions are meant to apply not just to scientists, but to people in general. This is not to say that there are no interesting ramifications of these issues for the philosophy of science, but merely that the thrust here is more general. Whilst the core part of the discussion is thus kept largely independent of issues in the sciences, I shall make reference to important connections to them at various points throughout the thesis. These scattered remarks will be collected in the last chapter of the thesis, where they also receive a more extended discussion.

¹³ For examples of this, see Howson & Urbach 1993 and Earman 1992.

2. Towards a Compelling Theory of the Representation and Alteration of Beliefs:

The Plan of the Thesis

In general, the argument of the thesis proceeds in four stages: firstly, the historical and conceptual *origins* of the probabilistic representation of beliefs are discussed. This lays the ground for the second step, a critical overview of the classic versions of these representations in the literature, together with an inquiry into the way in which these representations can be said to be “rational”. This account of the probabilistic representation of beliefs is then used in the third step as the basis of an inquiry into the justificatory base of Bayesian Conditionalisation as a means of altering beliefs. Finally, I present a new justification for the rule in the fourth and final step. In more detail, this suggests the following plan of the thesis:

Firstly, the historical and conceptual origins of the Subjective Theory of Probability (the general account of the representation of beliefs of importance here) are presented in chapter II, in the form of a critical inquiry into the early Keynes' and the early Carnap's "Logical Theories" of probability. These theories are seen to ultimately flounder on their inability to account sufficiently well for numerical measures of degrees of (rational) belief and (partial) entailment. Their only way of assigning numerical values to these is by relying on some version of the "Principle of Indifference", which turns out to be so flawed as to be inadmissible into any compelling theory of probability. However, out of this failure can be seen to spring the two phoenixes of modern probability theory: the

Frequency Theory and the Subjective Theory, both of which can be interpreted as – very different – reactions to the difficulties encountered by the logical theories.¹⁴

On this basis, a critical reading of the Subjective Theories that currently have the most currency becomes possible, which is tackled in chapter III. Firstly, a critical understanding of Ramsey's classic "Truth and Probability" provides the necessary basis on which to theorise further: as it contains many of the core issues at the heart of the present thesis, starting the inquiry with this essay makes the exposition in the later stages easier.

After that, de Finetti's (early) work is presented as a different take on the subject, one that is however broadly in line with Ramsey's. The crucial point of difference that is important here is in de Finetti's use of the "Dutch book argument" to justify the axioms of the probability calculus. This type of argument brings out clearly how important the distinction between logical and pragmatic rationality is, and it is also greatly relevant for the discussion in chapters IV and V.

The critical overview of the representation of beliefs then moves onto an extended look at the classic treatments of Savage 1954 and Jeffrey 1983. The former generally counts as the standard treatment of subjective probability and decision theory, so that a clear understanding of its achievements and limitations is crucial for the investigation here. On the other hand, a critical exposition of Jeffrey's 1983 treatment is important

¹⁴ The former though will not be discussed in the present thesis, as it is not directly related to the representation of beliefs. Also, note that the aim of the present inquiry is not a detailed treatment of the philosophy of probability, but deals with the latter only in so far as it is relevant for questions of belief alteration and representation. See also chapter III.

since it strengthens the conceptual foundations of Savage 1954 and links it to modern work in standard deductive logic.¹⁵

The investigation of the theoretical representation of beliefs concludes with a closer look at a key modern treatment of this topic: that of Howson & Urbach 1993. In the present circumstances, the two most important aspects of this work are, firstly, the way in which Howson & Urbach come to a theoretical representation of beliefs, and secondly, their criticism of the methodology of classical statistics. The latter is mainly important here for the way in which it reveals some of the consequences of the adoption of an account of the representation (and alteration) of beliefs for the way in which science proceeds.

I conclude this chapter with an assessment as to where all of this leaves the representation of beliefs by means of probabilities. Here, it turns out to be the case that there are strong arguments as to why it is pragmatically rational to have beliefs that are representable by means of probabilities. With a view to the logical rationality of his representation, the picture is more mixed, but there are some signs that a positive answer could be achieved in due time here, too.

Having thus concluded the inquiry into the probabilistic *representation* of beliefs, the discussion moves on to an extended look at the *revision* of beliefs by means of Bayesian Conditionalisation. In chapter IV, I present a proof of the soundness of this rule and discuss the two core arguments currently in the literature that aim to justify its importance. This discussion shows that these justifications provide no grounds for seeing the use Bayesian Conditionalisation as a matter of pragmatic rationality.

¹⁵ It also presents a different treatment of belief revision, together with some criticisms of Bayesian Conditionalisation. As pointed out above, this aspect of his work is less important in this thesis, however; some references to it are nonetheless made at various points throughout the thesis.

I therefore present a novel account of such a justification in chapter V. It presents an argument for the pragmatic rationality of the rule that is based on its ability to coherently structure an agent's dealings with the world and her fellow agents. I also present a case study that brings out some of the pitfalls of this novel justification, and connect it to Jeffrey Conditionalisation as a rival rule for the alteration of beliefs.

Lastly, I conclude in chapter VI with an overview of what the thesis has established, also in connection with some current and traditional issues in the philosophy of science. Finally, I present an outlook towards further work that could and should be done in this area.

3. Summary

On the whole, this work conducts its investigation into the probabilistic representation and alteration of beliefs (both in everyday life and science) by first presenting a closer look at the classic treatments of the probabilistic representation of beliefs and secondly by an investigation into the logical and pragmatic rationality of Bayesian Conditionalisation as a way of revising beliefs.

It is hoped that in this way the present thesis will not only throw some light on the difficulties and opportunities of finding a convincing theoretical account of the representation and alteration of beliefs, but that it will also go a little bit both towards illuminating the logical and pragmatic structure of decision theory and making sense of what can be expected of a modern philosophy of science. In order to do this, it is best to begin by laying the ground for further discussion with a closer look at the conceptual and

historical origins of the probabilistic representation and alteration of beliefs: the Logical Theory of Probability.

II. Historical and Conceptual Origins of the Subjective Theory of Probability: The Logical Theories of Keynes and Carnap

A very useful preliminary step in the development of a probabilistic representation of beliefs is a brief overview of the origins of the debate surrounding it. These origins are found in the attempts to come to a better understanding of what the notion of “probability” is about. One of the earliest philosophically rigorous formulations of the conceptual underpinnings of this notion is the “Logical Theory of Probability”. Despite the fact that it has to count as largely eclipsed¹⁶ by the various versions of the “Subjective Theory” presented in the next chapter, it still deserves a closer look here: gaining an appreciation of where its difficulties lie will give the subsequent discussion a much firmer conceptual and historical grounding.

The version of the theory that is the main focus of this chapter is that set out by J. M. Keynes, as it is the one that places the greatest weight on the relevant philosophical aspects, and because of its thorough discussion of the “Principle of Indifference”. However, references will also be made to Carnap 1950 as an alternative version of the theory (where this is deemed relevant).

The aspect of the theory I will concentrate on most is the abovementioned Principle of Indifference, which forms an integral part of this theory – both in the version propounded by Keynes as well as in that of Carnap. The latter does not appeal to the principle

¹⁶ That said, the recent upsurge of interest in “Objective Bayesianism” (an upshot of the *Subjective Theory* – laid out in chapter III – that constrains subjective degrees of belief further, normally by appeal to a maximum entropy principle) shows that some of the key ideas of the Logical Theory remain worthwhile points of exploration. However, even though these Objective Bayesians also see probabilities as representing *rational* degrees of belief, they would still have to maintain that the Logical Theory *as it is presented here* has to be considered eclipsed by the modern developments in the subject. Key references to Objective Bayesianism are Williamson 2005 and Williamson 2006.

directly, but arguably has to rely on something that is structurally similar in his choice of c-function (that is, whether it is to be based on state or structure descriptions).¹⁷ However, as will be clear below, this principle leads to grave difficulties in the form of logical paradoxes and has to be given up.

To show how these difficulties ultimately make a move away from the Logical and towards the Subjective Theory necessary, I firstly set out the contours of the Logical Theory and anchor the Principle of Indifference therein. Secondly, in parts 2 and 3, the paradoxes of the principle together with attempts to solve them are presented and evaluated. Lastly, I show how this discussion leads to a natural development away from the Logical Theory of Probability and towards the Subjective Theories as pioneered by Ramsey.

1. The Logical Theory and the Principle of Indifference

Keynes defines probability in two ways, which he takes to be equivalent: as “degree of partial entailment” and as “degree of rational belief”. Seeing these two as equivalent, however, already is contentious, both inside and outside the framework of the Logical Theory: Popper disagreed with Keynes and argued that they need not coincide (Popper 1999, p. 363-364) whilst Carnap held that degree of rational belief is essentially an empty concept that – analytically – must be equivalent to partial entailment (Carnap 1950, §41-§51).

The former concept (degree of rational belief) refers to the fact that, unlike in (valid) deductive inferences, the conclusion of an inductive or probabilistic inference need not be

¹⁷ See Carnap 1950, especially §46-§48 and section 4 below.

true if all the premises are, i.e. it is not *logically entailed* by them. However, according to Keynes and Carnap, it is at least sometimes possible to express the *degree* to which a conclusion is entailed by the premises (see for example Carnap 1950, §41 and Keynes 1921, p. 3-4 & p.11) – that is, to specify the *extent* to which the premises imply the conclusion (for valid deductions, this is the maximum amount possible). In general, this degree is being simply “known” by someone contemplating these propositions – it is given by intuition.¹⁸

“Degree of rational belief” refers to the “appropriate” cognitive stance justified by the propositions¹⁹ in question, i.e. the strength of belief of the quintessentially rational being. This emphasis on an ideal reasoner is important for what is to follow, as it is a strand of thinking that runs through much of the debate of the representation and alteration of beliefs. Roughly speaking, this rational degree of belief is being “revealed” to the *rational agent* through her intuitions (Keynes 1921, p. 18). These intuitions need not be as fully or as strongly present in *ordinary human beings* – they might make mistakes in terms of their strength of credence in some proposition, or fail to have any sort of intuitions on the matter. The explanation for this though has to be sought in the limitations of ordinary human beings, and not in the Logical Theory itself, which mainly concerns itself with rational agents (Keynes 1921, p. 18).

The three aspects of this characterisation of probability of the greatest importance here are the following: firstly, it implies that there really is no sense of talking about the

¹⁸ Keynes often uses the Russellian expression “direct knowledge” for this: Keynes 1921, p. 4 and especially p. 52. It is however important to note that Jeffrey draws a distinction between this kind of probabilistic intuition, and his own (and to some extent Ramsey’s) “radical probabilism”: see Jeffrey 1985, p. 66.

¹⁹ Note that Keynes is clear in stating that “propositions” are the basic terms of his theory: Keynes 1921, p. 4. In chapter III, I argue that this terminological choice can be seen to show some of the underlying motivations of the theory. Here, though, the emphasis is on a different set of issues so that there is no need to go into detail about this.

“probability of H” – probabilities can only be assessed on the basis of some evidence or some set of premises (i.e. an argumentative structure). In this way, one can really only talk about the “probability of H *given* evidence E”, so that the phrase “the probability of H” must be understood as shorthand for “the probability of H given the commonly accepted body of evidence” (Keynes 1921, p. 6 and Carnap 1950, §31).

Secondly, this definition of probability tries to strike a balance between the two “Janus faces of probability”: the subjective and the objective.²⁰ In general, the Logical Theory is one of the subjective theories, since it stresses the importance of “belief” and is not based on objective mathematical frequencies (as Von Mises’ theory for example).²¹ However, unlike in purely subjective theories (on which more below), there is only one “right” way of assessing probabilities - it is degree of *rational* belief, and it is easily possible that some actual agent’s opinion of the probability of some proposition is “wrong”. It is an aim of the present chapter to show that these two strands of thinking about probability cannot be so easily combined, and that their tension ultimately causes the Logical Theory to break asunder. This calls for a complete reevaluation of the notions of belief and entailment used here, and for a closer look at the distinction between logical and pragmatic rationality.

Thirdly, there is nothing in the above definition of probability that guarantees that it is measurable. In fact, as Keynes 1921 argues, there are many circumstances where only “ordinal” comparisons are possible – and some situations where even those are

²⁰ See Hacking 1975, especially chapters 1 and 2, and Gillies 2000, chapter 1.

²¹ Although it is important to note that this characterisation is somewhat contentious: Carnap 1950 argues that the Logical Theory - and, even more interestingly, the Subjective Theory, too – are as objective as the Frequency Theory: see Carnap 1950, §42-§46.

impossible.²² This latter will be the case when the probabilities are being based on different sets of propositions as evidence. This problem of the *measurement* of probabilities is also crucially important for what follows below.

However, numerical representations of probabilities *are* possible if the “Principle of Indifference” can be applied to the case at issue (Keynes 1921, p. 41), thereby aligning the Logical Theory with the mathematical accounts of the subject. This principle goes back to much older attempts at a formulation of probability, and versions of it can be found in Leibniz, Laplace and Bernoulli (often under the heading of “the principle of non-sufficient reason”).²³ The principle can most easily be stated as follows (Keynes 1921, p. 41-42):

Principle of Indifference: As long as there are no known reasons to assume that the propositions in question have different probabilities (given the evidence) one should proceed by assuming that their probabilities are equal.

The classic example is that of a coin: if there are no good reasons to assume that the coin is biased to a particular side, the probability of heads is $\frac{1}{2}$.²⁴

2. Problems and Paradoxes of the Principle of Indifference

Despite the seemingly obvious cogency of the principle, it is deeply flawed: it quickly leads to logical paradoxes. Here, I set out three particularly striking ones to illustrate the

²² See Keynes 1921, chapter 3, and particularly p. 38-39.

²³ See also Hacking 1975, chapters 14-15

²⁴ This also holds if it is only unknown on *which* side the bias falls.

general gist of these difficulties. These illustrations are then used to show how these difficulties form the basis for the inception of the Subjective Theory. Note please that this section is not a comprehensive taxonomy of all the possible paradoxes, but merely sets out the three core ones for purposes of illustration.

The first paradox is known as the “book paradox”²⁵: suppose we are speculating about the colour of the cover of a particular book. Assuming that this colour is completely unknown to us and that we have no reason to consider one of the colours more likely, the Principle of Indifference applies. We can therefore use it to deduce that the probability of the proposition that the cover is red (given the evidence) is equal to the other option, that the cover is non-red, that is, $\frac{1}{2}$. Of course the same reasoning could be used for all colours, so that the probability of yellow is also $\frac{1}{2}$, and so on. Then, however, we arrive at the conclusion that the sum of the probabilities of all possible colours (a mutually exclusive set) is more than 1. This contradicts the addition law of probability – and common sense.²⁶

Secondly, consider the “wine/water paradox”.²⁷ Assume there is a mixture of wine and water in a certain glass, and that the mixture contains at least as much water as it does wine and at most twice as much water as wine. The Principle of Indifference applies and can be used to deduce that the ratio water to wine has a uniform distribution over the interval 1 to 2. In particular, the probability that the mixture has a value in the interval [1, 1.5] should be the same (and equal to $\frac{1}{2}$) as that of the interval [1.5, 2]. Now however, consider the inverse ratio, wine to water. By the same reasoning as above, one would

²⁵ Formulations of it can be found in Keynes 1921, p. 43 and Gillies 2000, p. 37-38.

²⁶ A similar example can be created using the proposition “Person X is an inhabitant of Great Britain”; a separate discussion of this paradox seems unnecessary. See also Keynes 1921, p. 44.

²⁷ I follow here the version given by von Mises 1981, p. 77; other versions are in Keynes 1921, p. 45 and Gillies 2000, p. 38.

deduce that a value in the interval $[0.5, 0.75]$ and $[0.75, 1]$ has the same probability of $\frac{1}{2}$. However, these two situations are incompatible: The wine/water ratio of 0.75 corresponds to the water/wine ratio of $\frac{4}{3}$, implying that the range be split there at (roughly) 1.33 and not 1.5, which is a contradiction.

The third paradox leaves the realm of thought experiments and moves into the sphere of geometry.²⁸ Consider a circle with the inscribed equilateral triangle ABC. We are asked to give the probability that the randomly selected chord XY is longer than the side of the equilateral triangle. There are three main ways to argue about this problem:

Firstly, note that it is a geometrical fact that the centres of the sides of the triangle ABC lie at the midpoint of the radius r of the circle (see figure 1). Note further that the chord XY is greater than the side of ABC if and only if it lies between the side of triangle and the midpoint O of the circle. In other words, the chord is greater than the side of ABC if and only if the distance of the chord's centre W and the midpoint O of the circle is less than $r/2$. As the chord was randomly selected, the Principle of Indifference applies and can be used to argue that every value in the interval $[0, r]$ is equiprobable. Distances in the interval $[0, r/2]$ imply that XY is greater than ABC, which implies that the probability of this state of affairs obtaining is $\frac{1}{2}$ (one half of the range of possible values represent "favourable" instances).

The second argument considers the tangent to the circle at A and the chord XY with $X=A$ (see figure 2). Note here the geometrical fact that the angle α of the triangle is 60° and that therefore the chord XY is larger than AB if and only if the angle β between the chord and the tangent is in the interval $[60, 120]$. By the Principle of Indifference, any

²⁸ I follow here Gillies 2000, p. 38-42 most closely (I have however made some slight changes to the presentation of the figures to make the exposition clearer); similar versions can be found in Keynes 1921, p. 47-49.

particular value for β is equally likely in the interval $[0, 180]$. Since the interval $[60, 120]$ is a one third of the extent of $[0, 180]$, this means that the probability of the length of the chord exceeding the side of the triangle ABC is $1/3$.

Lastly, consider the circle with radius $r/2$ inscribed in the larger circle as above (see figure 3). By reasoning parallel to that in argument 1, it is clear that XY is greater than AB if and only if the centre W of XY is inside the area of the smaller circle. Again, by the Principle of Indifference and by the fact that the chord was randomly selected, point W can be anywhere within the main circle. Out of the total area thus “available” to W (πr^2), the area of the small circle ($\pi(r/2)^2$) represents favourable instances. This gives a probability of the chord being greater than AB of $(\pi r^2)/(\pi(r/2)^2)=1/4$.

Figure 1

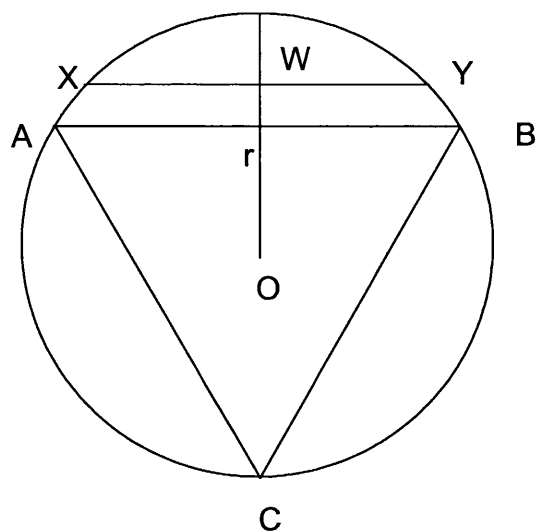


Figure 2

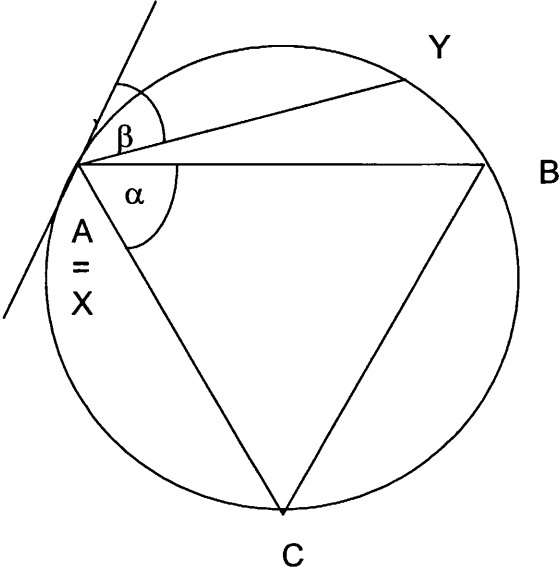
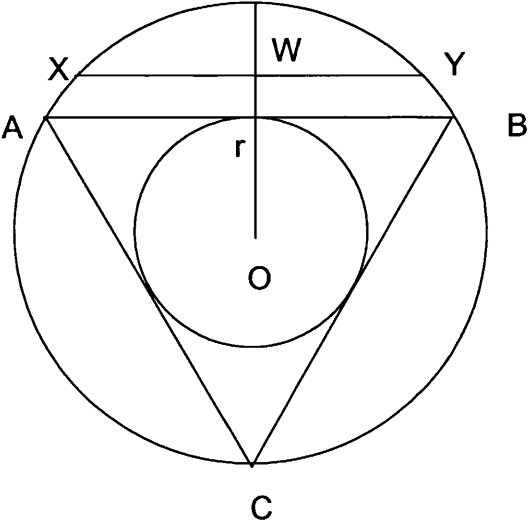


Figure 3



Hence, this paradox results from the fact that the same probability can be assigned different values by the Principle of Indifference depending on the particular argument chosen. In fact, this reasoning can be generalised by noting that the second and the third paradoxes were created by (correctly) using the Principle of Indifference multiple times,

once to argue that the values of parameter in question are equiprobable and then once or multiple times to some suitably defined continuous function of these values, arguing this time that the values of the function of the parameter are equiprobable. Normally, however, these two probabilities will differ.²⁹

This is a grave problem for the Principle of Indifference, as its primary purpose was to give unique numerical values to probabilities where it applies. The paradoxes however show that the Principle of Indifference alone cannot fulfil this role: there are cases where it is legitimately applied but does *not* specify a unique probability value – it is not clear which of these is the “right” one (if there is one at all: maybe all of the values are “wrong”).

In conclusion, it remains to remark that the Principle of Indifference does not provide the Logical Theory with the necessary methods to arrive at a numerical representation of even the restricted class of cases where it applies. It would therefore seem that the principle as set out above is untenable in the face of these paradoxes.

3. Solutions to the Paradoxes and their Success

Multiple solutions have been proposed to try to save the Principle of Indifference from refutation through these paradoxes. I will discuss and evaluate the two that have the most bearing for the issues of the present inquiry.

²⁹ This is true unless all values of the function are equal to the parameter values, i.e. $f(x)=x$.

3.1. Keynes' Solution

Keynes' own solution is to block the derivation of the paradoxes by constraining the applicability of the principle (Keynes 1921, p. 59-64):

Principle of Indifference (modified): As long as there are no known reasons to assume that the events in question have different probabilities (given the evidence), *and given that the events are weakly indivisible*, one should proceed by assuming that their probabilities equal.

“Weak Indivisibility” (my terminology) refers to the situation where one cannot subdivide the alternatives further into a series of disjuncts, each of which is similar in kind to the original alternative and which are mutually exclusive. In the case of the book paradox, this at first appears to block the derivation of the paradox, as the new constraint on the Principle of Indifference does not allow its application: “not-red” is divisible into the further alternatives “yellow”, “blue” etc., which are mutually exclusive and of the same kind as “red”.

It is however important to note that this solution relies on a particular language: for example, if our language were based on the colour spectrum, then “white” and “red” would not be of the same kind, the former comprising the latter (Howson & Urbach 1993, p. 61-62). Thus, in cases like this (involving what could be construed as *discrete* random variables) Keynes has to assume a particular language with a certain set of atomic propositions. There is no guarantee that the solution will give the same results for different languages (Howson & Urbach 1993, p. 61-62). Moreover, this reliance on

atomic propositions creates the familiar metaphysical difficulties of a world full of atomic propositions (whose ontological status is difficult to assess). These are problems that will also come up in the discussion surrounding Ramsey 1926.

At first, this modification does not seem to say much about the other two paradoxes: Since they involve a continuous interval, i.e. an infinite amount of alternatives all equal in kind, it does not seem clear how to apply the principle when the events in question are infinite. Keynes argues however that one can make sense of this by considering the limit of these infinite series. Take the example of a line of a given length: subdividing it into ever smaller intervals leads to an infinite amount of these intervals, each containing a set of points. The probability of a randomly selected point falling into any of these intervals is by the Principle of Indifference one over the total number of alternatives. Keynes' trick is to argue that that as the amount of intervals grows arbitrarily large, their size grows arbitrarily small, but will remain finite. This could be construed as a basic definition of a probability density function: the probability of the point falling into any particular interval is practically zero, but this does not mean that the size of these intervals is literally zero: it is only so in the limit.

Applied to the geometrical paradoxes this implies that the chord in question should not be seen as literally a chord, but merely as the approximation of the limit of a specific area as it approaches zero (Keynes 1921, p. 63). Keynes argues that since the areas in question are different, the limits are so too. Hence it is no surprise that the calculated probabilities differ: it only *looks* as though we are calculating the same quantity three times over, whilst in reality we are actually calculating three *different* probabilities.

However, it is not clear that this solution is enough to save the Principle of Indifference completely. There are two reasons for this. Firstly, it is very dubious whether one can really claim that the chord in question is nothing but the approximation of the limit of a specific area – this does not seem to accord well with common geometrical practice, where a line is not normally so defined. Moreover, Keynes merely states that because the areas differ, their behaviour towards the limit differs as well. As such, this might be contended: to see this, note that the original problem arose because there was a clash between the demand that there ought to be only *one* rational degree of belief in what intuitively seem to be identical situations, and the fact that the Principle of Indifference yields *multiple* ones in these situations. Keynes’s solution was to modify the Principle of Indifference in such a way that the situations under discussion are no longer identical: they might concern the same limit, but differ in their passage towards this limit. However, it seems that we still need a justification for *this* fact – we might also want to hold on to the intuitively plausible position that what matters for rational degree of belief is the *limit* of a continuous function, not the passage towards it. If so, we would be willing to reject any principle – like Keynes’ – that has other consequences. In a nutshell: one might dig one’s heels in to say that the probabilities *ought not to differ*, no matter what Keynes’ modifications actually entail. It would then be up to him to argue for his principle on other grounds; and that he has not done.

Secondly and more importantly, Keynes’ solution remains mute as to the wine/water paradox. This still goes through, even in the modified version of the Principle of Indifference: it is difficult to argue that the “shape” of the water/wine ratio is different to that of the wine/water ratio.

I therefore conclude that Keynes did not present a valid solution to these paradoxes: he did not address all of them and even those he did address cannot be considered adequately solved.

3. 2. Jaynes' Solution

A second set of solutions has been proposed by Jaynes 1973. In these, Jaynes focuses on the third paradox and proposes a constraint on the number of permitted transformations of the original parameter set in order to make sure “artificial” derivations of paradoxes can be avoided (e.g. using the reciprocal of some quantity instead of that quantity directly, for example in the case of height). To be exact, he introduces certain “invariance principles” which must be satisfied to license the application of the Principle of Indifference. For example, in the case of the geometric paradoxes above the only argument that is rotation, scale and translation invariant³⁰ is the first one, which would therefore give the unique correct probability of the chord being greater than the side of the equilateral triangle as $\frac{1}{2}$.

However, yet again, this solution does not go far enough. Firstly, it still leaves the wine/water paradox untouched. None of the transformations proposed by Jaynes applies to this case and it is dubious how any other one could, given the symmetry of the two ratios in question. Secondly, one could argue that these invariance principles are somewhat ad hoc: In effect, one would have to formulate a different set of them for every possible paradox and moreover, it is not at all clear why this particular set should be the only “correct” one. What is so special about rotation, scale and translation invariance for example? If the goal is simply to reduce the number of alternative solution arguments to

³⁰ Actually, one really only needs translation invariance: Jaynes 1973, p. 485

one, then it will not seem convincing to claim that the probability thus arrived at is really the “right” one. It will merely show that it is possible to weed out various solutions until a unique one remains. In principle, it would still seem possible to get different “unique” probabilities by using different invariance principles. This can hardly be considered an adequate solution to the initial problem.³¹

Jaynes however proposes a third argument for the Principle of Indifference: its practical success. He claims that the principle has been used often and with much success throughout much of the history of science, especially in physics. Therefore, the argument goes, the alleged paradoxes simply cannot be that important or damaging to the applicability of the principle (see Jaynes 1973, p. 479 and Gillies 2000, p. 47).

However, this defence also seems to fall flat: the objections presented in section 3 were meant to show that the Principle of Indifference often fails to yield unique values for probability statements. Jaynes’ argument from success shows that the principle can *sometimes* be successfully applied in practice, something that the objections did not deny. This seems to be a confusion of quantifiers: The objections argue that $(\neg \forall x)(Tx)$ whilst Jaynes argues that $(\exists x)(Tx)$ ³², which do not stand in logical opposition. Gillies 2000 frames this problem as one between heuristic and logic: the Principle of Indifference might be a useful heuristic device, but just like any other heuristic device does not yield logical truths. The probabilities derived from it had been and must be tested empirically just like any other scientific theory (Gillies 2000, p. 48).³³ However, what the Principle of Indifference set out to do was not to give some rough practical guidelines suggestive of

³¹ For a class of related problems, see also the discussion of Jeffrey 1983 in chapter III below.

³² Where Tx could be interpreted as “The Principle of Indifference yields a unique probability for the circumstance x .”

³³ This though is a point that Jaynes 1973 readily agrees with, e.g. p. 480 and p. 485.

further empirical work, but rather some way of deriving unique numerical values of probabilities at least in some cases. This last has been cast in doubt by the paradoxes (especially with a view to the language-relativity of the Principle of Indifference), not the former, which again shows the shortcomings of Jaynes' solution.

This last point can be restated fruitfully by drawing on the two different senses of "rationality" distinguished in chapter I. At most, what Jaynes could be arguing for is that scientists are *pragmatically* rational in trusting the Principle of Indifference. That is, he might claim that trusting the principle tends to have beneficial results for the scientists using it. However, even arguing in this more careful manner is unconvincing, and that for two reasons.

Firstly, it is still the case that it is not *logically* rational to trust the principle: as the paradoxes above show, the Principle of Indifference has to be discarded as a logical principle, since it is inconsistent. Given the fact that the Logical Theory is a framework where this kind of rationality would seem to matter greatly (for example, it is concerned with ideal reasoners only, it is about partial entailment relations holding between propositions etc.), this is a major shortcoming.

Secondly, Jaynes' claim that using the Principle of Indifference actually *is* pragmatically rational might be called into doubt as well. Truly, it *might* be the case that the situations that give rise to the paradoxes are simply not ones that physicists will ever have to face. Given the generality of the underlying structure of the paradoxes, however, this might appear somewhat dubious. Whilst this is admittedly a question for the experts in the field (i.e. the physicists, which might also include Jaynes himself), there do not seem to be any clear a priori grounds for why it *should* be pragmatically rational to use

the principle. At the very least, Jaynes does not give any. Short of this, all that Jaynes' argument has shown is that the Principle of Indifference has actually yielded some good results in the past. This, however, is a very different claim from one stating that the principle also represents a strong *tendency* to do so.

Generally, therefore, no satisfactory solution to the paradoxes has been suggested – no solution, that is, that blocks the derivation of *all* the paradoxes; in fact, the second paradox seems to go through in all of the proposed solutions. Moreover, even in those cases where some of the paradoxes are being blocked, the solutions do not seem convincing (e.g. due to mistaken or unconvincing claims concerning its pragmatic and logical rationality). I therefore conclude by noting that the Principle of Indifference must be considered refuted.

4. From the Logical to the Subjective Theories

These considerations concerning the Principle of Indifference have immediate consequences for the Logical Theory of Probability. In order to see this, it is best to go back and reconsider why the Principle of Indifference was introduced in the first place: to give a numerical representation of probability. The paradoxes and the inconclusive solution attempts have shown that it fails in this task.

Two options now seem open: either one accepts this conclusion and tries to save the Logical Theory by making no use of numerical considerations, or one keeps the desideratum of a numerical representation and adapts the Logical Theory accordingly.

Using Lakatosian terminology, one would in the first case define the “hard core” of the theory as the fact that probability is defined as degree of rational belief and partial entailment; it might then be considered merely an unfortunate fact of reality that these probabilities are inherently immeasurable.

However, this leaves one with the awkward consequence of a having a fully fleshed out mathematical theory of probability (the logical axiomatisation of which was at the heart of Keynes’ project) that is really not “about” anything. It merely gives a purely counterfactual account of how one could deal mathematically with probability if one were able to give a numerical representation of it (which one is not). Alternatively, the mathematical calculus could be seen as describing a mathematical concept that has no counterpart in reality, so that Keynes’ treatise is really about two different concepts, both confusingly called “probability”. Neither construal seems very convincing, however. Many, if not all, sciences rely on numerical probability considerations, so that these moves would turn the Logical Theory into a highly degenerating research programme.³⁴

It is thus much more promising to find another answer to the irrefutability of the paradoxes of the Principle of Indifference. On this view, one takes the numerical representation of probability to be part of the hard core of the theory and tries to modify the theory to retain this element. Since the Principle of Indifference failed as a method for numerical representation, an alternative way of doing so has to be found.

This alternative way becomes clearly visible if one takes a step back to note that one of the key difficulties of the Logical Theory is its attempt to hold on to probabilities being *unique*, whilst at the same time wanting to construe them as having to do with *degrees of*

³⁴ Note here Lakatos’ scathing criticisms of such a research programme in Lakatos 1978, for example, p. 134-135.

belief. Prying these two apart immediately offers two ways of providing a numerically measurable notion of probability. This can be seen most easily by noting that the conception of the uniqueness of probability in the Logical Theory could be seen to be both too strong and too weak.

It is too weak if probability is seen as an objective feature of the world (of the sort of “mass” and “volume”), in which case it seems greatly misguided to take its unique determination to lie merely in the intuitions concerning the degrees of belief of an ideally rational agent. Rather, it is much better to drop talk of degrees of belief altogether in this context. One would thus concentrate on the element of the Logical Theory which assigns one correct probability to every proposition and drop the Principle of Indifference to instead use mathematical generalisations of empirical regularities (of sorts) to arrive at the desired numerical representation. This is the step taken by von Mises in his Frequency Theory (see for example von Mises 1981).³⁵

On the other hand, the Logical Theory’s sense of uniqueness is too strong if one takes the fact that probability has to do with an agent’s degrees of belief seriously. On a subjectivist construal, what a theory of probability should provide are requirements of consistency or coherence across degrees of belief (see also Skyrms 1986, p. 130). From this angle, the dilemma that the Logical Theory faces is the following: either it relies on the Principle of Indifference and suffers from the paradoxes, or it drops the Principle of Indifference and suffers from having mysteriously intuited, non-measurable “probabilities”. This dilemma can be overcome by finding a different way of probabilistically representing beliefs that is based neither on mysterious intuitions, nor on

³⁵ This, though, will not be discussed further in the present context, as it is not at the centre of the present inquiry. For more on it, see Gillies 2000, Howson and Urbach 1993, chapter 13 and Jeffrey 1977.

the Principle of Indifference – effectively the step taken by Ramsey and the other subjectivists (as shown in chapter III), and the core of the approach taken by modern Objective Bayesians.³⁶ Note also that this does not rely on taking the agent to be a real agent – even ideal Bayesian reasoners can be free (at least within pre-defined limits) to assign degrees of belief as they wish to a set of propositions.

The consequences of these “modifications” are obvious – they leave little of the Logical Theory and effectively create new theories of their own. Positively speaking, however, they make clear that in some sense the Logical Theory could be considered as the foundational theory on which the other theories grow. This also justifies why it is still useful to examine it rather closely even today: other interpretations of the philosophy of probability really only come to life with it in the background. It thereby sets the stage for many of the important discussions surrounding the Subjective Theory (and belief revision rules), for example the notion of “rationality” that is being appealed to and the problem of the measurement of probabilities.

Interestingly, many of the points raised by Keynes are picked up and given a radically different treatment just five years after the publication of his “Treatise on Probability” by Frank P. Ramsey in his classic “Truth and Probability” of 1926. This essay forms the basis of all modern treatments of the representation and alteration of beliefs in a probabilistic setting and is therefore the focus of the first part of the next chapter.

³⁶ Interestingly, Carnap 1950 seems to opt for both options at the same time: he uses “betting quotients” – an operationalist way of measuring degrees of belief (see also chapter III) – without however specifying why there is only one correct betting quotient in every situation (as he claims). At the same time though, he also relies heavily on the Frequency Theory (his probability₂) in defining the Logical Theory (see also Carnap 1950, §41). It is also important, however, to be aware of the fact that over the course of his life, Carnap changed his mind and became a much stronger proponent of the Subjective Theory than he was when writing the “Logical Foundations of Probability” in 1950. However, I shall not pursue these historical issues any further here.

III. The Probabilistic Representation of Beliefs: An Overview

The inquiry into the logic and pragmatics of a theory of the representation of beliefs now reaches its second stage: given the overview of the conceptual origins of such a theory sketched in the previous chapter, it is now possible to move towards actually *finding* a compelling theoretical representation of an agent's beliefs. Such a representation is clearly necessary in order to continue the inquiry into a theoretical account of the *alteration* of beliefs: alteration of beliefs can only be made sense of in a theoretical framework if some way of *representing* them had already been found. This then is the task of this chapter: to give a secure foundation on which to begin scrutiny of the rule for updating beliefs given in chapters IV and V.

The way I proceed in this is by presenting the main treatments of this problem in the literature, in what is essentially a chronological order (though it will also turn out to be very convenient from a conceptual point of view). This not only allows me to make clear what the current state of the debate is in this sector and where it comes from, but it also already throws up many of the issues that will be returned to in the discussion of the alteration of beliefs. The classic treatments I shall look at in some detail are Ramsey 1926, de Finetti 1937, Savage 1954 and Jeffrey 1983. As a representative of a more recent take on these issues, I shall concentrate on Howson & Urbach 1993. The aim of these discussions is to make clear the difficulties and opportunities that the probabilistic representation of beliefs faces and how these have been approached to date. The foundation and in many ways still the most important and influential of all of these theories is Ramsey's, with which I shall begin.

1. Ramsey's "Truth and Probability"

The first and – as it will turn out – conceptually most important reaction to the failure of the Logical Theory to provide the kind of rigorous foundations for a theory of probability is due to the English mathematician, logician, economist and philosopher Frank P. Ramsey. His method – though historically largely overlooked and misunderstood – stands as one of the most impressive intellectual achievements in the whole history of 20th Century analytic philosophy.

Here, I present a critical statement of the core tenets of Ramsey's 1926 approach. This will prove to be extremely helpful in the further investigation of the representation and alteration of beliefs that follows later, as much of the debate surrounding it is foreshadowed in Ramsey's classic. Also, since many of the later treatments can be seen as continuations and expansions of Ramsey's work, it is important to be clear about his ideas from the outset.

1.1. Ramsey's Method

The core idea behind Ramsey's take on the question of the theoretical representation of beliefs is the following: an agent's subjective degrees of belief in the truth of some proposition are measurable on a probability scale using her preferences over various choices. The first point to note about this is that there is a slight shift in the direction of the argument here: whereas Keynes and Carnap were interested in exploring the conceptual foundations of a certain mathematical construct – "probability" – and found that it could be construed as having to do with the beliefs of an ideal reasoner, Ramsey

comes at the topic from the opposite direction. He starts by inquiring into “the logic of partial belief” (Ramsey 1926, p. 69), and then finds that in fact, subjective degrees of belief can be construed as probabilities. In a certain sense, he therefore clearly also gives a new, subjective, foundation to the theory of probability (subsequently termed the “Subjective Theory”), but it makes his argument much clearer if this is seen as the consequence, and not the starting point, of the method. It is also vital for the present thesis, which, following this approach of Ramsey’s, sees probabilities merely as instruments for a successful theory of beliefs, but not as objects of study in their own right.³⁷

Given this way of arguing, Ramsey has to make two related points: firstly, he has to give a method of measuring degrees of belief numerically. Secondly, he has to show that this way of measuring degrees of belief satisfies the axioms of the probability calculus.³⁸

If he were to fail in the first task – i.e. if it were to turn out that degrees of beliefs are not always numerically measurable – his theory would be much less useful in science and everyday life and might even have to face exactly the same kinds of problems that beset the Logical Theory of the previous chapter.

Failing in the second task is not so much a difficulty for Ramsey due to the direction of the argument stated above. Had it turned out that degrees of belief can only be given a different (i.e. non-probabilistic) theoretical representation, so be it. In fact though, it did not. However, it still has to be an element of his argument to show exactly why it is that degrees of belief turn out to be probabilities.

³⁷ Note please that I am not saying that it is not worthwhile to inquire into the conceptual foundations of probability, but merely that this is not the aim of the inquiry here.

³⁸ In some formulation of them. I shall not distinguish here amongst these various formulations, the canonical of which is Kolmogorov 1956.

Starting then with the first part, Ramsey employs what seems to be a strongly operationalist and behaviourist methodology.³⁹ On the operationalist side, for Ramsey a concept like "degree of belief" can only be considered to have content if there is some way of *measuring* that content (Ramsey 1926, p. 69). Expressed more strongly: the way of measuring the content of the concept just *is* that concept, i.e. measurement and concept are coextensive (Ramsey 1926, p. 70; Bradley 2001, p. 269&271).⁴⁰

In line with his behaviourism, Ramsey furthermore does not want to rely on the agent's subjective feeling of the strength of her belief to measure degrees of belief with. This, he argues, is bound to lead to mis-measurement, as some strongly held beliefs are not accompanied by any strong feelings at all (Ramsey 1926, p. 71). Moreover, measuring degrees of belief according to their strength of feeling might not lead to a scale that is anything like that of the mathematical theory of probability (Ramsey 1926, p. 73): that is, the agent might have an ordering of least noticeable differences of strengths of belief that are not isomorphic to the real numbers between 0 and 1.⁴¹

Instead, Ramsey suggests taking the individual's *behaviour* to establish their degrees of belief (Ramsey 1926, p. 72-73). That is, he proceeds by measuring degrees of belief through their causal properties on the agent's actions (Ramsey 1926, p. 71&73). In a move that is echoed in some of the work of Davidson (especially Davidson 1963),

³⁹ This sort of methodology was quite within the spirit of his time, which saw the rise of behaviourist psychology and was just before the onset of the philosophical behaviourism of the positivists; for some remarks concerning it, see Bradley 2001, p. 265-266 and Gillies 2000, p. 100.

⁴⁰ Ramsey is not fully clear as to which of these versions of operationalism he actually endorses; fortunately, for present purposes this distinction is not so relevant – see also Bradley 2001, p. 269.

⁴¹ This is not actually the argument that Ramsey makes. Instead, he argues that the scale of strength of belief might be different from the scale of probabilities, for example by having more points in the interval $[0, 1/2]$ than in $[1/2, 1]$ (Ramsey 1926, p. 73). However, this is merely a problem of scaling the respective measures. Since there is only one agent in the theory, this is unproblematic. The version of his argument given above is thus somewhat of a generalisation of Ramsey's actual point (or, less charitably, it is the argument that Ramsey *should have* made, but did not). I thank Mauro Rossi for helpful discussions of this point.

Ramsey here argues that an agent's desires and beliefs *cause* her to act in certain ways. Thus, given suitably chosen actions, an individual's *choices* over them can be used to measure her *degrees of belief*. Ramsey argues that the traditional suchlike actions are bets, and that this method is "fundamentally sound" (Ramsey 1926, p. 73). He however moves on to generalise this situation by not considering bets as such, but an individual's choices over "conditional prospects" – in essence, possible worlds described with reference to antecedent conditions: "if this coin comes up heads, we will go have Italian food for dinner, if it comes up tails, we shall have Thai food" (see also Ramsey 1926, p. 76-77).⁴²

Before one takes a closer look at these conditional prospects and the framework in which they are in, it is important to note that Ramsey does not claim that the only evidence his measurement theory rests on must be behavioural in the sense of being based purely on observation of actual choices. Rather, he is basing his method on "hypothetical choices" – on reports of how the agent *would act* if he did face a choice situation of the specified kind (Ramsey 1926, p. 72). He is thus allowing *some* introspection into his system (namely reports of the introspected behaviour in hypothetical choices) so that his clear behaviourist stance should not be conflated with an all-out *anti-introspective* position (see also Bradley 2001, p. 287).

Given this particular behaviourist and operationalist approach, Ramsey proceeds as follows. He firstly introduces the notion of an "ethically neutral proposition", a proposition the truth or falsity of which is of no importance to the agent – for example, the weight of some rock at the bottom of the ocean (Ramsey 1926, p. 77). These ethically

⁴² These conditional prospects are often referred to as "gambles" in the literature on Ramsey (for example, Jeffrey 1968, p. 30-31), but I shall follow Bradley 2001, p. 270 and solely use Ramsey's own terminology of 'prospects' here.

neutral propositions are crucial for Ramsey, as they function like an anchor for his measurement procedure. Intuitively, this can be made sense of as follows.

Since (as stated above) Ramsey assumes an agent's actions are *caused* by her desires and beliefs (Ramsey 1926, p. 74), the aim of his method has to be to find a way of splitting the causal origins of an action into its cognitive and desire-related components. In general, this is not straightforward, as these components have to be identified simultaneously. This means that in order to get the system off the ground in the first place, a third element has to be found that can be used to pry them apart; this element is the "ethically neutral proposition".⁴³ By basing his theory on these, Ramsey can disregard the desire component of the action (as the individual is completely indifferent as to their truth⁴⁴) and use them to present the agent with a choice situation where she only acts according to her *beliefs* ("uncontaminated" by any desires). Given this, it becomes possible to recover the degree of belief in an ethically neutral proposition from choices over conditional prospects that involve them. In this way, Ramsey can proceed to "normalise" the degree of belief in ethically neutral propositions by assigning it a value of $\frac{1}{2}$ (Ramsey 1926, p. 77).

Building on this, Ramsey uses an agent's preferences over conditional prospects together with the ethically neutral propositions to come to a measurement of degrees of belief in any proposition (ethically neutral or not) as follows. He firstly assumes that the desirability of a conditional prospect is given by the weighted sum of the desirabilities of

⁴³ Here (as elsewhere), the influence of Ramsey on Davidson becomes clearly visible: see for example Davidson 1976, p. 270 and Davidson 1974, p. 145-147.

⁴⁴ Note an ambiguity in the formulation here: for Ramsey, desires and preferences attach to (conditional) prospects, and not to propositions (Ramsey 1926, p. 77-78). This means he has to spell out more clearly how the ethically neutral propositions relate to prospects, a difficulty that will be commented on in slightly more detail below (see also Bradley 2001, p. 271&283).

the prospects (i.e. the possible worlds) involved, where the weights are given by the agent's degrees of belief in the prospects' occurrence (Ramsey 1926, p. 75). This is the core part of his method – without it, his whole take on the measurement of beliefs would collapse. Crucially, this take on how desires over conditional prospects are constructed is an *assumption* of Ramsey's method: it is not being explained by it.

Secondly, this is enough for Ramsey to set up what is essentially an algebraic difference structure in order to measure the agent's degrees of belief as a probability (see Ramsey 1926, p. 78-80).⁴⁵ The main idea behind this is that given the agent's desirability ranking of the various prospects involved (Italian food, Thai food etc.) and the previously defined conditional prospects (which lie in between the various simple prospects due to the nature of their desirability in the form of an arithmetical average), a mapping can be found that links all the degrees of belief in the (conditional) prospects to the real-valued interval $[0, 1]$. Also, a function (which is unique up to positive linear transformations) can be found that measures her 'utilities' over these prospects (Ramsey 1926, p. 78-80).

Slightly less abstractly, the procedure can be illustrated as follows. Assume N is an ethically neutral proposition; by the above arguments, Ramsey can assign it a probability value of $\frac{1}{2}$. Now, consider some agent who turns out to be indifferent between getting prospect A if N is true and B if it is false, and prospect C with certainty. Then it follows immediately that the utility of C must lie on the midpoint between A and B : $u(C) = \frac{1}{2}u(A) + \frac{1}{2}u(B)$ (by assumption). This process can be repeated (with more complex conditional prospects) to obtain the desired *utility* function.⁴⁶

⁴⁵ The technical details of this are not greatly relevant in the present context; for a closer treatment of them, see Bradley 2001, p. 276-278.

⁴⁶ The non-uniqueness of this function derives from the fact that the zero-point and the unit interval can be chosen arbitrarily: see also Eells 1982, p. 68-69.

This utility function can be used, again with the assumption that the utility of any conditional prospect is given by its mathematical expectation, in order to obtain the *probability* measure for any proposition. To see this, assume that the agent is indifferent between the conditional prospect “getting prospect A if Q is true, and B if not”, and getting prospect C with certainty. Then it follows by similar reasoning to the one above that

$$u(C) = \text{prob}(Q) u(A) + (1 - \text{prob}(Q)) u(B)$$

and therefore

$$\text{prob}(Q) = \frac{u(C) - u(B)}{u(A) - u(B)}$$

This can be repeated for any proposition to obtain a total probability function as desired. For what is to follow, it is important to note that this function *is* unique: zero-point and unit are fixed and cannot be changed at will (this will become important in the discussions of Savage 1954 and Jeffrey 1983 below).⁴⁷

It is finally useful to also draw attention to one last aspect of Ramsey’s method: the alteration of beliefs. He argues that an agent should base her degrees of belief after learning a new piece of evidence on the conditional degrees of belief she had previously – i.e. setting $\text{prob}_{t+1}(H) = \text{prob}_t(H/E)$ (Ramsey 1926, p. 87). This is interesting for two reasons: firstly, it represents a very early and clear statement of the conditionalisation rule that will form the centrepiece of the fourth and fifth chapters of this thesis. Secondly, the brevity of Ramsey’s remarks here betrays the immense intuitive appeal of the principle of conditionalisation that he had formulated. It appears entirely obvious to him that updating one’s beliefs in this way is the only rational option and does not even stop to consider

⁴⁷ The above derivation largely follows Eells 1982, p. 69.

whether there might be problems with this principle. These are points I return to in chapters IV and V, so I shall leave them aside for now.

On the whole, therefore, Ramsey's method works by positing desires and beliefs as the fundamental causes of actions. One particular kind of action – choices over (conditional) prospects – can then be used to measure degrees of belief by setting up an algebraic difference structure that together with the existence of an ethically neutral proposition is sufficient to prove that degrees of belief satisfy the axioms of the probability calculus, i.e. can be measured as probabilities. In this way, Ramsey has fulfilled the two tasks he set for himself above.

1.2. Criticisms and Replies

At this stage it is very helpful to consider various objections to the approach taken by Ramsey. This not only makes the philosophical arguments that he is grappling with much clearer, but also highlights some difficulties that are very important for the further development of the theoretical representation of beliefs and their conditionalisation in the later parts of this thesis.

Firstly, there are some technical difficulties in the notion of Ramsey's ethically neutral proposition. As Bradley 2001 (p. 274) shows, in order to spell out clearly within his system of prospects (or possible worlds) what an ethically neutral proposition is, Ramsey ends up being committed to the early Wittgenstein's theory of atomic propositions (or "facts" and "states of the world") to a much larger extent than might be visible on the surface. This is problematic, as reliance on these atomic facts brings with it well-known metaphysical problems: for example, whether atomic propositions correspond to atomic

facts (creating a bloated ontology) or if negated atomic propositions have a negative truthbearer etc.⁴⁸

It is not clear to what extent Ramsey can fully deal with this problem. Bradley 2001 (p. 288) suggests a move towards a more Jeffrey-inspired account just in terms of propositions (i.e. not involving prospects at all), but leaves the details somewhat unspecified. Sobel 1988 (p. 237) on the other hand claims that Ramsey is not committed to the above propositional atomism at all. Be that as it may, this is not a point of major relevance for present purposes. It concerns a rather technical detail, and given the general project of the representation of beliefs pursued here, the costs (if any) with which Ramsey has to repair a somewhat minor technical glitch are not greatly relevant.

A graver issue is the extent to which operationalism and behaviourism are crucial cornerstones of Ramsey's method. Concerning the former, the difficulty is this: as Gillies 2000 (p. 138-145) has argued, operationalism of the kind espoused by Ramsey seems very much an outdated approach to the philosophy of science. Ever since the decline in popularity of the kind of logical positivism that Ramsey's operationalism is a precursor of, philosophers and scientists are far less worried about theoretical terms that cannot be directly measured. Clearly, measurement remains vitally important for any empiricist scientist, but this does no longer mean that measurement procedure and the concept to be measured have to be as closely related as operationalists make it out to be.

Fortunately, Ramsey does not depend on an operationalist philosophy very much and could drop most of it without a single change in his method. Instead of claiming that he

⁴⁸ A closer discussion of these difficulties is neither necessary nor possible in the present context; the two classic treatments of logical atomism are Wittgenstein 1921 and Russell 1918.

wants to set degree of belief *equal* to its measurement procedure, he would then merely claim to present *one way* of measuring the former (see also Bradley 2001, p. 269).

The problem concerning Ramsey's *behaviourism* on the other hand might be more taxing on his method. The worry here is not so much that it is too extreme – as argued above, he only rules out certain kinds of introspective evidence – but that it is not *clear* how far it goes: on closer inspection, "hypothetical choices" turn out to be far more complex than it might appear at first sight.

The main trouble with hypothetical choices is that asking someone how they *would* choose seems not too different from asking them what they *feel* or *believe* – how else would they decide how they would decide? That though means that, really, Ramsey is relying on exactly the kinds of judgement that he wanted to avoid: subjective assessments of the "strength of belief". In this way, his detour through the causal theory of intentional action comes to naught, since *hypothetical* choices do not turn out to be choices at all.

These difficulties moreover gain plausibility from the fact that Ramsey's reply to them falls far short of being convincing. It appears that he wants to argue that there might be a subconscious common cause of both action and mental state, and that it is this common cause that the method is seeking to estimate (Ramsey 1926, p. 72). He argues further that it therefore does not matter whether the method is based fundamentally on the mental state or the action, since they all reduce to the same underlying phenomenon anyway.

However, this reply does not seem to provide a strong defence of his method, since it is very mysterious what that common cause could be, if not itself some sort of mental state. In this case, though, it seems that Ramsey is in fact falling back onto the

methodology of simply asking the subject about the workings of her mind – the very approach he rejected earlier.

However, there is a much more cogent response to this difficulty that might capture the point that Ramsey was after, but failed to make clear. This response consists in shifting the argument from considerations of the metaphysics of the mind to the epistemology of Ramsey's system. The crucial idea here is that subjects might find it much easier to *respond* to a question about how they would choose as opposed to one about how strongly they feel. This might be a contingent fact about the method, and not a necessary tenet of its foundations. As was already stated above, Ramsey is not completely averse to introspection, so that in favouring hypothetical choices over straightforward introspection of "feelings", he might be merely be drawing attention to the epistemological advantages of one method over another, without making any principled claim about what it is in the mind that causes these difference (see also Bradley 2001, p. 266). In this way the reliance on hypothetical choices opens up a whole new range of behavioural *evidence* – speech behaviour – that otherwise would have remained absent (see also Bradley 2001, p. 287).

The third criticism is due to Jeffrey, who claims that Ramsey's method forces the agent to assign strange and bizarre causal powers to the person eliciting her beliefs (Jeffrey 1983, p.157). Imagine Jones asking Smith whether she would prefer a cup of coffee with certainty to becoming Prime Minister if a (fair) coin lands heads, and subsistence farmer in rural mainland China if it lands tails. Given such a choice, Smith either has to take Jones for having some mysterious power that can make all of these things happen at the toss of coin (Prime Ministers normally are elected by the general

public, and being a farmer requires skills, land and capital)⁴⁹ or she has to assume that Jones is in fact lying (because the outcome of the coin toss just does not bring about these situations). In either case, Smith would order her preferences in the light of this extra knowledge, so that her choice over these prospects would be no guide whatsoever to her *true* preferences simpliciter: they would already be conditioned on one of the two hypotheses above. That is, she – at best – could be seen as revealing what her preferences over these gambles are *if she is faced with an omnipotent agent or a liar* – and these might be very different from her preferences tout court (see also Jeffrey 1968, p. 33).

However, in Ramsey's defence, it is important to keep in mind here the previous point about what Ramsey's behaviourism actually amounts to, and in particular, his affinities to *hypothetical* choices. That is, Smith in the above example only has to answer Jones' question *as if* she thought Jones could bring about election results or fundamental changes in space-time, but not by actually *believing* it. That is, Ramsey merely asks the agent to consider how she *would* choose if she had to face a choice of this sort.⁵⁰ This seems to be not so out of line from what could be expected of an agent in such a situation (see also Bradley 2001, p. 287).⁵¹

The most important objection that might be raised against Ramsey's approach however is that its core principle – namely that the desirability of a conditional prospect is given by its mathematical expectation – is entirely unjustified.⁵² Of course, this early version of

⁴⁹ Indeed, Ramsey claims that Jones has to persuade Smith that she has the "power of the almighty" (Ramsey 1926, p. 76).

⁵⁰ This is also another interpretation of the "power of the almighty" passage cited above: Jones only has to *persuade* Smith that she has the powers of the almighty, not to *actually* have them (Ramsey 1926, p. 76).

⁵¹ Clearly, there are many well-known difficulties with "as if" approaches (for an amusing account of one of them, see Worrall 1989, p. 267-268). However, these are problems that do not attach to Ramsey's method any differently than they do to the general case, so that I shall not discuss them further here.

⁵² Ramsey merely calls it a "law of psychology" that he "introduces" (Ramsey 1926, p. 75); Bradley 2001, p. 270 also sees it as completely unjustified.

expected utility maximisation had been known for a while, at least when it came to betting behaviour (it features for example in Bernoulli's argument against the St Petersburg paradox – see Joyce 1999, p. 34). In terms of a compelling justification of it, however, more is needed than that: there are many hypotheses that are well known, *yet false*. At the very least, what is necessary is an argument that aims to show that this particular principle of human psychology is (normatively or descriptively) convincing. However, despite the fact that Ramsey needs this principle for his method to be workable at all, he gives no real grounds for believing in it. This seems to put the entire system on very weak foundations: for how can the theory be considered compelling if its crucial column is given no argumentative support?

As a defence of Ramsey, one might argue with Bradley 2001 (p. 263-265) that it is important to distinguish two different projects: on the one hand, Ramsey might be seen to want to show that it is a requirement of rationality⁵³ to take the desirability of a conditional prospect to be given by its mathematical expectation and to have degrees of belief that fall in line with this principle. On the other hand, he might be seen to take a much more limited approach and be only concerned with the *measurement* of degrees of belief, independently of their justification.

One could then argue further that it is in fact this latter project that Ramsey is subscribing to: he takes it for granted that another theory has to be given why the agent should be assumed to be an "expected utility maximiser". This though runs perpendicular to his own project, which concerns the different – but equally important – issue of the

⁵³ Where this last term would have to be further specified as being either of the pragmatic or the logical sort.

measurement of degrees of belief, given some other theory of the above sort (see also Bradley 2001, p. 264).

However, whilst at first compelling, it is important to keep in mind the historical and conceptual situation that Ramsey is in, also with a view to the present thesis. At this point, we do not have any such supporting theory – there is no “standard decision theory” to fall back on, for the simple reason that this is exactly what is to be developed. The problem at the outset of the investigation was to find a compelling way of theoretically representing beliefs. If Ramsey can only do this with the help of some further theory that provides the normative support for his core principle, then his method cannot be considered adequate unless and until he has developed this further theory. As was established above, he has done nothing of this sort.

This also the reason why the distinction between the two projects above does not really help Ramsey’s case: the project of measurement piggybacks on that of providing a normative foundation for the theory (see also Bradley 2001, p. 264). However, at this point in the inquiry, this other project has not gotten off the ground yet. Up to now, Ramsey has merely shown that *if* an agent were to apply mathematical laws of expectation to conditional prospects as he assumes, *then* her degrees of belief are measurable by his method. This is an important result for many different fields (in fact, it can be seen to have started the whole area of measurement theory), but it leaves much to be explained when it comes to provide a compelling theory for the representation and

alteration of beliefs. In fact, it is exactly a theory of this latter type that is needed to satisfy the antecedent of the conditional claim above.⁵⁴

1.3. Conclusion

On the whole therefore, Ramsey presents a fairly comprehensive and convincing method for the measurement of beliefs. Its main problem is that it lacks justificatory foundation – it does not give any grounds why people should obey the kind of principles it posits. This means that what is most needed at this point is to find a justification for the particular way (i.e. using the mathematical expectation) in which an agent is meant to assign a desirability value to objects like the conditional prospects that play such a crucial role in Ramsey’s system. One of the earliest and most intuitive suchlike justifications is based on the idea that violating this kind of “expected-utility maximisation” makes the agent open to guaranteed losses of money. Exactly this is the route taken by de Finetti.

⁵⁴ The only way that Ramsey’s project could be salvaged in the present context is therefore by firstly developing a decision theory that is independent of his method, and then returning to the latter for the measurement issue. However, I shall not pursue this complication, for the simple reason that most decision theories do not seem in need of an independent measurement procedure (as argued below).

2. De Finetti's "Foresight: Its Logical Laws, Its Subjective Sources"⁵⁵

De Finetti's take on the theoretical representation of beliefs is not only notable for its originality, but also for presenting an approach to the topic that is still at the core of many of its modern treatments: that based on betting. This on its own would be enough to warrant a sufficiently detailed treatment of his views; what makes it all the more pressing is that the approach he has initiated is – despite its popularity – very controversial. In order to avoid misconceptions and needless arguments later on, it is therefore vital to have a solid grounding in de Finetti's ideas: many of the problems unearthed here are relevant in the later parts of this inquiry and can be studied most easily in their original form as presented by de Finetti. I begin by presenting his method, and then go on to discuss the main objections that can be raised against it.

2.1. Dutch Books and the De Finetti Theorem

De Finetti approaches the theoretical representation of beliefs (just like Ramsey) in an operationalist and behaviourist manner (de Finetti 1937, p. 101).⁵⁶ However, unlike Ramsey, de Finetti has a much more narrow conception of the behaviourist evidence allowed: the only choice situation that is considered adequate are *actual* bets (de Finetti 1937, p. 101). This makes him differ from Ramsey quite considerably, who only speaks

⁵⁵ A brief historical note is in order here on the relation between de Finetti and Ramsey: Ramsey read "Truth and Probability" to a distinguished audience of scholars at Cambridge in 1926. The paper itself got published in 1931, 6 years before de Finetti published his "Foresight: Its Logical Laws, Its Subjective Sources" (in 1937). However, de Finetti did not learn about Ramsey's paper until much later in his career, and over time came to accept Ramsey's utility-based methodology as superior: see Gillies 2000, p. 50-51 & 56-58 and de Finetti 1937, p. 102 (note (a)).

⁵⁶ For a good overview over some of the issues connected to these two positions in de Finetti's philosophy, see Galavotti 1989, especially p. 240-241.

of conditional prospects⁵⁷, and also presents the most controversial and most original element of de Finetti's approach.

Before I go into more detail concerning his exact method, it is worthwhile to note that he constructs his theory around the term "events" (de Finetti 1937, p. 102) – a noticeable difference to Ramsey's "propositions" and "prospects". It also marks the beginning of a division in the subject between those theorists who prefer to talk about "events" (like de Finetti and Savage) and those who instead choose to employ terms like "propositions" that are more directly of linguistic nature (as for example Ramsey and Jeffrey, and later on Howson & Urbach). As will become clear below, this is largely a terminological issue, but does *point* to a fundamental difference as to whether these theories see themselves as situated more on a pragmatic or more on a logical plane.

To return to the details of de Finetti's method: he proposes to use the following betting scheme to elicit an agent's degrees of belief in the occurrence of some event.

Firstly (as Ramsey did before him) de Finetti assumes an agent's beliefs in conjunction with her desires to be causally efficacious in bringing about her actions (de Finetti 1937, p. 101). The particular kind of desire that is in question for de Finetti is the desire for money (de Finetti 1937, p. 101-102). *Ceteris paribus*, an agent desires more money to less. For simplicity, it is also assumed that the agent does not care – i.e. is indifferent about – the variability of the money that is at stake. That is, she does not care about the riskiness of the gamble: for her, only the expected value of the bet matters, not its variance. This is an important and very strong assumption that will be discussed in more detail below. For now, it is best taken for granted.

⁵⁷ As was argued in the previous section, these should not be over-easily conflated with simple bets, despite Ramsey's insistence that his approach "is based fundamentally on betting" (Ramsey 1926, p. 81). The difference in the extent to which de Finetti and Ramsey rely on betting behaviour is made clear below.

In the second step, the agent is offered the following bet: she is to choose a real number q (her “betting quotient”) in the knowledge that she will get $(S - qS) = (1-q)S$ in case the event in question occurs, and $-qS$ otherwise. The final twist in the tale is that her interlocutor gets to choose the real number S (de Finetti 1937, p. 103; see also Gillies 2000, p. 55).

Given this setup, it can now be easily shown that unless the agent sets her betting quotients such that they obey the axioms of the probability calculus, a clever bookie can always make money off her – i.e. there is the possibility of a Dutch book being made against her. This can be shown as follows.⁵⁸

Assume (without loss of generality) that the agent bets with betting quotients q_1 to q_n and stakes S_1 to S_n on the occurrence of a set of mutually exclusive and exhaustive events E_1 to E_n . Then:

1. Assume that $n = 1$. Now, if $q < 0$, the bookie is sure to win by setting $S < 0$. On the other hand, if $q > 1$, then the bookie is sure to win by setting $S > 0$. Moreover if E is the certain event, then if $q < 1$, the bookie is sure to win by setting $S < 0$.

Thus, $q \in [0, 1]$, and $q_E = 1$ for the certain event E .

2. Assume now that $n > 1$. Then the gains (or losses) to the agent can be spelled out in the following matrix, where row i corresponds to the i th event occurring (remembering that only one of the n events can occur, but that one of them must):

$$\begin{array}{cccc}
 1 - q_1 & -q_2 & \dots & -q_n \\
 -q_1 & 1 - q_2 & \dots & -q_n \\
 \dots & \dots & \dots & \dots \\
 -q_1 & \dots & \dots & 1 - q_n
 \end{array}$$

⁵⁸ Here, I shall follow mainly the approach as it is actually set out in de Finetti 1937, p. 103-110; other versions can be found in Gillies 2000, p. 60-64, Howson & Urbach 1993, p. 79-81 and Earman 1992, p. 39-40.

This system will have no solution – so that there is no sure gain to be made by the bookie – if and only if it is singular, i.e. its determinant is zero. This determinant is given by $(1 - q_1 - q_2 - q_3 - \dots - q_n) = 1 - (q_1 + q_2 + \dots + q_n)$.

Thus: for a set of n mutually exclusive and exhaustive events:

$$\sum_{i=1}^n q_i = 1$$

3. Assume the situation is as above, only that the following definition of a “conditional bet” is introduced (de Finetti 1937, p. 109): a bet on E_1 conditional on E_2 is a bet on E_1 that is being called off if E_2 does not occur (with all stakes being returned). Then consider the agent making following three bets: She bets with q_1 on the occurrence of $(E_1 \& E_2)$, with q_2 on the occurrence of E_2 and with q_c on the conditional bet (E_1/E_2) .

The relevant matrix is in this case (where the first row corresponds to the case of E_1 and E_2 both occurring, the second to only E_2 occurring and the third to neither E_1 nor E_2 occurring):

$$\begin{array}{ccc} 1 - q_1 & 1 - q_2 & 1 - q_c \\ -q_1 & 1 - q_2 & -q_c \\ -q_1 & -q_2 & 0 \end{array}$$

Once again, this system will have no solution if and only if the determinant is zero, i.e. if and only if $q_1 - q_2 q_c = 0$.

Thus: $q_1 = q_2 q_c$.

These three cases correspond to the three standard axioms of the probability calculus, so that it has been shown that an agent wanting to avoid the possibility of a Dutch book

being made against her in a betting scheme of the above sort will set betting quotients that satisfy the axioms of the probability calculus.

In a final step, the previous two results are combined as follows. By assumption, having a Dutch book made against her goes against the desires of the agent: since she prefers more money to less, she *will* seek to avoid this situation and set her betting quotients so that they *do* obey the axioms of probability. This allows de Finetti to identify these betting quotients with the agent's degrees of beliefs, since it was assumed that beliefs are causal, and as the desire element has already been accounted for (in the form of the desire to avoid unnecessary losses of money). This means that the only other element with an influence on the agent's choice behaviour that is left to be accounted for are her beliefs.

In this way, de Finetti comes to develop a measurement procedure that is based on the presumed desire of an agent to avoid Dutch books. From this, it immediately follows that the instrument with which degrees of belief are measured – betting quotients – must satisfy the axioms of the probability calculus. This is a surprising and groundbreaking result that deservedly has sparked a massive discussion up to today.

2.2. The Limits of the Betting Approach

There are two kinds of objections that could be and are commonly raised against de Finetti's method. The first set of them centres on the general gist of his methodology, and the second focuses on the particular nature of the betting approach that he envisions. I discuss each of them in turn.

The first set of worries centres on de Finetti's operationalism, and that for similar reasons as in the case of Ramsey. Once again, the theoretical construct "belief" is given content solely through its measurement procedure: the procedure *is* what beliefs *are* (de Finetti 1937, p. 148).⁵⁹ This has the awkward consequence – also noted in the context of Ramsey's approach – that if there is another consistent measurement procedure for beliefs, then they have to be seen as defining two different entities ("belief₁" and belief₂" maybe).⁶⁰

However, since these difficulties have already been discussed in the context of Ramsey's approach, I shall not pursue them any further here. Moreover, just as it was true there, it remains the case here that much – if not all – of de Finetti's ideas on the measurement and theoretical representation of beliefs can be translated without loss into a non-operationalist framework. This makes worries about de Finetti's general operationalism rather less pressing.⁶¹

However, the second set of problems is much more important. Instead of concentrating on the general philosophical framework of de Finetti's arguments, they centre around the particular structure of his approach, and in particular its reliance on "betting".

At this point, it is very helpful to note that de Finetti's reliance on the desire to avoid Dutch books is to be strongly contrasted with Ramsey's approach. The latter's argument might be construed as being based *structurally* on betting situations,⁶² but, crucially,

⁵⁹ De Finetti is clearer in defending the strong version of operationalism than is Ramsey: see also Galavotti 1989, p. 240-241.

⁶⁰ For more on this and other difficulties with operationalism, see also Gillies 2000, p. 138-140 and in particular p. 139.

⁶¹ For more on de Finetti's general operationalist outlook, see also Galavotti 1989, especially p. 239-243.

⁶² See also the quote given in footnote 57 above.

nothing he says relies on the possibility of Dutch books at all. For him, the fact that degrees of belief are measurable as probabilities has to do with the nature of rational preference (an account of which he presupposes). The situation is very different for de Finetti however; as was made clear above, his derivation of the possibility of a probabilistic representation of beliefs depends crucially on the desire to avoid guaranteed and unnecessary losses of money in a betting situation of the sort sketched out by him.⁶³ The second set of difficulties now attaches to exactly this kind of reliance on bets.

The gist of this set of difficulties is that the betting situation that de Finetti works with is highly specific: it involves paying qS for the chance of winning S if some event occurs (see also Howson & Urbach 1993, p. 78). Why should exactly this kind of bet be considered the one to be used in order to find a theoretical representation of beliefs? There seem to be many other kinds of bets and gambles (the variety of suchlike devices on offer in any casino bears witness to this) that – at least on the face of it – seem to have an equally good claim to be considered alongside de Finetti's approach.

However (and this is the main part of the objection) most of these will not yield anything like the theorem de Finetti has proven above. In most cases, it seems that these gambles either do not allow any sort of belief-measurement; in some others, the kind of measurement they provide is either very crude or provably non-probabilistic. If this is true, then the foundation for de Finetti's theorem is rather weak: it relies on a very specific choice situation, whose general applicability seems greatly questionable.

Two answers to this charge seem possible. Firstly, de Finetti could deny that other gambles do not have the property of providing probabilistic representations of degrees of

⁶³ In fact, this could be seen as his attempt to provide the justificatory element missing from Ramsey's theory.

belief. For example, despite appearances to the contrary, Adams 1964 has argued that even a bet like “I pay you K pounds if it rains tomorrow, and you pay me K pounds if it does not” might allow for a probabilistic representation of beliefs. However, it is unclear how compelling such a wide-ranging defence of all kinds of gambles can be (given the variety of them on offer) – it simply seems implausible to see *all* of these as probabilistic measures of degrees of belief (the complexities of modern financial markets might be a case in point here).

Secondly, de Finetti might counter the objection by noting that indeed, there are many other kinds of bets, but that this is entirely irrelevant to the question at issue. It does not matter how many betting situations do *not* have the consequences of allowing probabilistic representations of beliefs, as long as *one* does. Consider measuring the length of a wire: here, also, it is completely unimportant how one *cannot* measure its length (using one’s sense of smell, say) as long as there is one way of doing so (for example, by seeing how many standard meters can be put along side of it). Thus, de Finetti’s betting scheme is to be preferred to any other ones simply because it in fact does allow the measurement of beliefs.

However, there is a need to be careful here. It is not the case that all it takes for de Finetti’s representation theorem to go through is to have the agent bet in the manner prescribed above: the betting situation *alone* does not do the trick. One further has to assume much about the agent’s psychological, social and cultural situation – for example, that she truly does have the desire to avoid unnecessary and certain losses of money in the betting situation set out above.

This further (implicit) assumption though means that a stronger version of the above difficulty could be run as follows: it is conceivable – and might even be likely – that an agent is very much indifferent to losing money in these sorts of situations. The reasons for this are to be found in exactly the kind of psychological, social and cultural background assumptions needed for the above argument to go through.

Proceeding in reverse order, consider the cultural and social assumptions first. In order for his method to work, de Finetti has to assume that “betting” in general is well understood and is considered appropriate in the culture in question. This need not be the case however: some religions forbid gambling for example, and “betting” might not be something that is at all practiced in some agent’s culture. More to the point, her culture might allow betting *in general*, it just happens to be the case that *the kind of* bet described above is not at all common in her society, so that the agent has no idea whatsoever about how to choose in this situation. In these cases, de Finetti’s whole approach would have to be considered ludicrous to represent or measure such an agent’s beliefs. Even disregarding his operationalism (which would entail that the agent in question does not have any beliefs at all, which seems very wrong indeed), it still means that the betting scheme envisioned by de Finetti is too specific to be applicable in the general case.

Furthermore, there are major *psychological* reasons for why the betting scheme might fail rather frequently. Firstly, consider someone who really enjoys playing these sorts of games, for their own sake. Someone like this might not see the possibility of a Dutch book as an *unnecessary* loss of money at all: just like there is a certain loss of money involved with any purchase, the agent might simply see herself as buying the fun of betting with the possibility of a Dutch book (just like she would not worry about “losing”

money in return for a cinema ticket). That is, she does not play to win, but merely for its own sake.

Secondly, there might be the opposite case of someone who rejects – out of some principle – to engage in any sort of gambling behaviour or refuses to take part in the specific bets that de Finetti has described. In this case, de Finetti’s scheme is entirely incapable of finding out about her beliefs, as there is nowhere for the betting methodology to latch onto.⁶⁴ Moreover, in order for de Finetti’s scheme to have problems, it is not even necessary that the agent completely *rejects* any betting whatsoever. As was mentioned above, a key assumption of his approach is risk neutrality – that is, the agent is assumed to have constant marginal utility of money (so that her utility is linear in money). That though need not be the case; in fact, much recent empirical as well as theoretical work in economics is predicated on the presumption that agents are risk averse.⁶⁵ Assuming risk neutrality also runs into conceptual problems due to the St. Petersburg Paradox (see for example Jeffrey 1983, p. 151-154).

However, if the marginal utility of money is declining, agents might be very reluctant to bet, even though the expected value of the bet is positive. This is greatly problematic for de Finetti’s theory, as in this case, the agent’s behaviour in the face of a bet is not just due to her degrees of belief, but also to the extent of her risk aversion.⁶⁶ That is, her risk

⁶⁴ Once again, if de Finetti’s strict operationalism is to be taken seriously, then one might even have to conclude that such an agent has no beliefs at all: if the content of the concept of belief is given by its measurement, then if there is no measurement, there is no content.

⁶⁵ See for example Mas-Collel et al. 1995 p. 483-488 and Ramsey 1926, p. 73 & 76 for this.

⁶⁶ This could be measured by using for example “relative risk aversion” or “absolute risk aversion” measures; details about and differences between these measures, however, are not important in the present context; see Mas-Collel et al. 1995 p. 185-187 for more on this.

aversion works like a very strong “disturbing cause” in de Finetti’s experiment: it disrupts the process in question, and that to such an extent as to become almost useless.⁶⁷

In a certain sense, this is the correlate of the situation in Ramsey’s system: there, Ramsey simply assumed that the individual would take the utility of the kinds of conditional prospects he sketches to be their mathematical expectation. Similarly here, de Finetti simply assumes that the individual takes the utility of money of the bets he sets out to be based on their mathematical expectation. In both cases, this has the consequence of being easily called into doubt.

De Finetti and his supporters have not let the case rest at this point, however, but went on to formulate replies to the charges. Starting with the problem of risk aversion, they have argued that whilst all of the above might be true, it does not spell trouble for de Finetti’s betting scheme in the slightest.

The reason is that diminishing marginal utility might well approximate linear marginal utility for low enough money values. That is, as long as the stakes chosen are small enough, the problems of risk aversion mentioned above should not be too grave: an agent might not be very risk averse if the possible gains and losses are not very high (de Finetti 1937, p. 102, note (a) and Jeffrey 1956, p. 16). Of course, this means that one is now in danger of escaping the throes of Charybdis just to run into the arms of Scylla: if the stakes are too small, the agent is in danger of not wanting to “bother about trifles” (Ramsey 1926, p. 76). In that case, she might not pay careful attention to the situation and her attitudes towards it, so that once again, the results of the procedure cannot be taken to accurately represent the beliefs of the agent.

⁶⁷ This way of construing the situation is in the spirit of Mill 1844, p. 60.

Defenders of de Finetti's approach argue that there is a way of trading off between these two dangers that leaves the core of de Finetti's methodology intact. That is, they claim there is a value (or a range of values) of the stake of the bet that is such that it is low enough to avoid the problems of risk aversion but high enough to ensure the agent is thinking about the situation she is facing properly (see for example de Finetti 1937, p. 102, note (a), Gillies 2000, p. 56 and Jeffrey 1956, p. 16).

However, it is highly questionable whether there really is such an amount. It seems that if this amount existed, it should be specifiable. However, to date no estimate of it whatsoever has been published, the likely reason being of course that the amount is different for different agents (depending on their overall level of wealth and many other psychological factors). This means that a defender of de Finetti would have to develop an *agent-specific* methodology for the establishment of this value that is moreover *independent* from de Finetti's procedure for the measurement of degrees of belief.⁶⁸ It is very unclear what such a procedure might look like, and at least, none has been set out as yet.

However, even if such a procedure were to be specified, de Finetti's method would not be out of trouble, as the original problem of its overly particular nature would return with a vengeance. To be fully workable, de Finetti's betting scheme would not only have to be based on one particular kind of betting situation, but this betting situation has to involve a specific kind of stake that will vary from agent to agent.

The problem with this is that one might be very doubtful that this method really measures the same quantity in every agent; it rather looks as though something very

⁶⁸ This last condition is needed to avoid problems of circularity, since the value (or range of values) of risk neutral amounts of money for the agent would then be *used* to measure her degrees of belief.

different is measured in every individual case. As an analogy, consider the measurement of length again: it is as if one would have to use one kind of instrument (a ruler, say) to measure the length of a table, and an entirely different one to measure the length of a chair (sound waves, say).

In order to be able to claim that these two procedures really measure the same quantity, further theoretical arguments are needed – both about these procedures, and about the nature of the quantity to be measured.⁶⁹ What these arguments have to make clear is that the different measurement procedures tap into the same underlying structure of the quantity in question: in the example above, arguments to that extent might be that length is *defined* as one-dimensional extension, and that both procedures above can in some circumstances be used to come to an approximate representation of the extension of an object along one dimension.

To return to the case here: de Finetti equally has to provide further theoretical arguments of this kind – which he has *not* done – in order to show that the many different measurement procedures reduce to the same “underlying” concept of “belief”. What is necessary here is a fully developed theory of this concept – not just of its measurement or representation, but also about its nature. Such a theory would be a major advance in the philosophy of mind (and the cognitive sciences in general); it certainly cannot be taken for granted or left in the background. Without saying more about this explicitly, de Finetti’s scheme cannot be considered adequate.

Given this, de Finetti might seek recourse in the claim that the whole approach should not be mistaken for a simple recipe for empirical work. Instead, his method should be

⁶⁹ Moreover, if de Finetti’s operationalism were taken at face value (so that the concept “length” is really defined by its measurement procedure), then one would again have to conclude that there are two different concepts in play here: “table-length” and “chair-length”.

seen as a *theoretical* approach: it is meant to hold only in principle, for somewhat idealised agents in somewhat idealised circumstances. There need not be a straightforward direct relationship between these idealisations and reality. In other words, he might argue that he gives a *principled* way of measuring beliefs, which need not be straightforwardly applicable in real life.

However, if de Finetti is forced to give this answer, his victory would only be pyrrhic. If the aim of the inquiry is the ideal agent only, then it is unclear what all the talk of risk aversion and refusals to bet is necessary for; in fact, using “bets” might already be unnecessarily concrete. It would be much better to instead base the elicitation of beliefs on more abstract “judgements” of the agent (as Ramsey did and Howson & Urbach will do later).

Doing this though would have two problematic consequences: on the one hand, it would cause his approach to collapse into that of Ramsey or Howson & Urbach. This means that the main advantage of de Finetti’s treatment – that a clear (monetary) justification for the rationality assumptions underlying the measurement procedure – is getting lost (see also below for more on this).

On the other, it goes against de Finetti’s pragmatist approach to the topic, which aims to be useful in real life (Galavotti 1989, p. 241-242). This usefulness might be seen as the motivation for using “events” as the main terms on which the theory is built, for basing it on bets, and for using money instead of abstract “utility”-like judgements. The aim is to present a *pragmatically* useful theory, not one that is for ideal agents only. In this emphasis on pragmatics, it adds an element to Ramsey’s more logical approach that will become greatly important in what is to follow. It also hints at the important underlying

distinction between logical and pragmatic rationality: if de Finetti aims to provide a justification for the fact that logically rational ideal agents have beliefs that can be represented by means of probabilities, his method is too specific and tied too much to the workings of actual agents. If, on the other hand, he seeks to provide a justification based on pragmatic rationality, his scheme is only a partial success, as it can only be applied in a restricted setting: having beliefs that can be represented probabilistically will only tend to be beneficial for not overly risk averse agents who live in an environment that allows the kind of bets that de Finetti sets out.

The final criticism to be levelled against de Finetti's betting scheme concerns the fact that simply *offering* someone a bet may *change* her beliefs: in these cases, measurement of an agent's beliefs fails, since the very procedure of doing so alters what is to be measured (this is a point also forcefully made by Davidson 1976, p. 271). For example, a bookie's offering a bet on Liverpool winning the Premiership this season might lead the agent (who might not follow British football much) to suspect the bookie knows more about this situation than the agent does, and revise her beliefs accordingly (from a higher to a lower value, say). This shows how difficult it is for de Finetti's scheme to keep apart issues concerning the representation and the alteration of beliefs, a point that is also of importance with a view to the discussion of chapters IV and V.

In all, what remains therefore is the fact that supporting a Ramsey-style measurement theory with an account that justifies Ramsey's core principle by an appeal to *economic rationality* is simply unhelpful (see also Howson & Urbach 1993, p. 90-91). The desire to avoid Dutch books in a betting situation is just not enough to build a compelling normative theory of belief representation (see also Ramsey 1926, p. 73).

2.3. Conclusion

In summary, I have tried to argue that taking bets as the basis for the measurement of belief cannot – on its own – be considered fully compelling (despite the fact that it might prove to be helpful in some cases). There is more to having beliefs than a disposition to bet in a certain way, and other considerations have to be taken into account even in those cases where betting behaviour might be used to represent beliefs. Importantly, these issues are also highly relevant for some similar kinds of justifications for the *alterations* of beliefs (as argued below).

However, some positive points can be still be taken from this discussion. On the one hand, de Finetti has unified the vocabulary of the theory by talking about “events” only (instead of Ramsey’s conditional prospects and propositions). On the other, he has tried to bring the theory closer to the reality of actual agents. Both of these points are taken up by de Finetti's main American follower Leonard Savage some 20 years later.

3. Savage's "The Foundations of Statistics"

The most important and popular current treatment of the representation of beliefs and desires in a theoretic framework is Savage's "The Foundations of Statistics". It is also the most well-known and most widely discussed *decision theory* in the literature at the moment and is at the basis of most modern economics and statistics (see for example Howson & Urbach 2004, p. 8-9 & Mas-Colell et al. 1995, p. 205).

For the present context, its main importance derives from its unification of the methods of Ramsey and de Finetti into one coherent treatment. It does this by combining elements from a framework of actual agents with a representation of beliefs that is fundamentally based on the preferences and 'utility' of an ideal reasoner. Crucially also, it seeks to provide the kind of *normative* justification for the postulates a measurement theory like Ramsey 1926 has to rely on.⁷⁰ However, unlike de Finetti, it does not seek to do this by means of principles based on economic prudence, but on fully fleshed "rationality" postulates that any (rational) preference ordering should satisfy.

In order to bring out clearly how it achieves this unification, and to what extent it can be considered an adequate theory of the representation of beliefs, I firstly present a sketch of the core of Savage's account and then bring up some of the most important criticisms that have been raised against it.

3.1. Savage's Decision Theory

The first thing that has to be noted at this point is that the focus of the inquiry has shifted for Savage relative to that of Ramsey and de Finetti. The former two were

⁷⁰ As argued above.

concerned mainly with the elicitation of beliefs⁷¹ whilst Savage is out to present a “decision theory”⁷² – a theory that gives structure to and makes sense of an individual’s decisions about how to act.

This is an important change in the thrust of the argument made and introduces a whole new set of issues to be tackled; the most important of these (for present purposes) concern Savage’s assumption of a causal theory of intentional action, which takes an individual’s actions to be determined by her prior beliefs and desires (see for example Savage 1954, p. 105) – similarly to Ramsey and de Finetti. This means that since Savage wants to make sense of an individual’s practical rationality, he has to first find a way to theoretically grasp these desires and beliefs. Clearly, it is this respect of Savage’s theory that takes centre-stage here.

The three core notions of Savage’s decision theory are acts, states and consequences (Savage 1954, p. 8-14). Consider the following table:⁷³

	State 1 (speed limit: 30 mph)	State 2 (speed limit: 50 mph)
Act 1 (drive 30 mph)	Consequence 1 (no speeding ticket, be late)	Consequence 2 (no speeding ticket, be late)
Act 2 (drive 50 mph)	Consequence 3 (speeding ticket, be on time)	Consequence 4 (no speeding ticket, be on time)

“States” are complete descriptions of possible worlds that leave nothing of relevance unspecified (Savage 1954, p. 8). Thus, “the speed limit on this road is 50 mph” and “the speed limit on this road is 30 mph” might represent two possible worlds that are presumed identical in all relevant aspects, apart from the speed limit on this road. Note

⁷¹ Or, given the direction of the argument of the second chapter, they were concerned with finding a consistent and compelling subjective interpretation of the theory of probability.

⁷² For more on the appropriateness of this term, see below.

⁷³ This is largely based on Savage 1954, p. 14.

also that the condition that there be no *relevant* aspect undescribed is quite crucial: for example, this road being quite obviously a country road, there being a clearly marked construction site nearby, or most trivially, there being a sign stating the speed limit are surely relevant aspects of the world that need to be mentioned in the description of the state. “The speed limit on this road is 50 mph” and “There is a sign that the speed limit is 30 mph” just do not sit easily together as descriptions of the same possible world.⁷⁴

There are two important points to notice concerning these ‘states’. Firstly, the description of such a state cannot include any aspect that it is in the agent’s power to change: anything that the agent might have influence over should be part of Savage’s *acts*. The state is meant to be a way the world might be, with it either being the case that the world actually *is* like that – or it not being the case. There must not be anything within the agent’s power bring it about that one or the other obtains. This means that “Tomorrow is a sunny day” and “Tomorrow is a rainy day” are states, whereas “Me being on holiday tomorrow” or “Me not being on holiday tomorrow” are not, since I can presumably influence the latter, but not the former (Savage 1954, p. 25).⁷⁵

Secondly, the stress on the *description* of the world means that the theory becomes focused on the specific use-interest in question. The same worlds might be described in many different ways and therefore ultimately call for very different actions (Savage 1954, p. 9). The crucial point to note here is that, as long as nothing of relevance is left out of

⁷⁴ Note that they do not *logically* exclude each other – relevancy here must be taken with reference to the conditions of the possible world in question. There is an epistemological difficulty here about the extent to which this relevancy can be easily specified, but in order to avoid unnecessary complications, I assume that this specification can indeed be done here.

⁷⁵ There is a worry here as to whether it is actually possible to meaningfully separate out all the elements under the agent’s control from the description of the state of the world, and still retain a description of the world that can be considered of interest to the agent. I shall not however discuss this point any further here, so as to avoid unnecessary complications.

the description, these differences are entirely in line with what one would expect when one has to deliberate about very different acts. Savage's theory fits this insight.

To make this clearer, consider the following example: when deciding whether to read "The Tempest" or not, one will be interested in very different possible circumstances than when one wonders whether to go ski jumping. For example, it might be better to describe the two states one takes into consideration in one's deliberations as "There will be a power-failure later on today" and "There will not be a power-failure later on today" in the former case, whereas in the latter, they might be more helpfully described as "There will be a snowstorm later on today" and "There will be no snowstorm later on today". However, these differences in description need not mean that these are actually four different worlds (as opposed to two).

One way to avoid this relativity is through taking as large a world as possible (that is, to take the entire history of the universe as one world with a completely specified past and future). The downside of this would be that this might cause significant complications that might be avoided using smaller worlds (see also Savage 1954, p. 9).⁷⁶ Fortunately, for present purposes, these issues are not very important, so I shall not debate them further here.

Finally, it is also useful to note that Savage combines sets of states into "events" (Savage 1954, p. 10-12), so that smaller worlds correspond to events in the larger world. Despite the fact that these 'events' correspond fairly neatly to the concept of a "proposition", it is better to follow Savage here and stick to the terminology of 'events': this provides a much better fit to his emphasis on the pragmatics of actual decision

⁷⁶ Note also that there are some problems concerning the relationship between big and small worlds; for a useful overview over these issues, see Howson & Urbach 2004, p. 10. I shall however not discuss this point further here.

making, as opposed to the logic of ideal decision making (see also Howson & Urbach 2004, p. 9). Terminological issues do become important if they make clear the *emphasis* of the theory, despite the fact that they themselves need not entail any actual differences in the theories' fundamentals.⁷⁷

Savage's second key concept is the "consequence" (the boxes in the table above): the combination of the agent's action and the state of the world that obtains (Savage 1954, p. 13-14). That is, the consequences list the outcomes of the agent's actions in the various possible states of the world. Ontologically speaking, these consequences will also be states of the world: they are once again complete (with respect to all relevant aspects) descriptions of the way the world is after the agent has acted in a certain way and a certain state of the world obtains (Savage 1954, p. 14).⁷⁸

The third and final crucial element of Savage's theory is the already mentioned "act" (the rows of the above table). These are the options for acting open to the agent in the given situation (Savage 1954, p. 14-17). Formally, they are *functions* from states to conditions (Savage 1954, p. 16):

$$a(s) = c$$

where *a* is an act, *s* a state and *c* a consequence. Given this functional description of acts, it immediately becomes clear that an act must specify a consequence for every state

⁷⁷ Note the similar issue in the context of de Finetti 1937.

⁷⁸ Here the independence of the states of the world and the agent's actions becomes visibly important; in fact, one could call 'consequences' 'states of the agent' to distinguish them from the non-agent 'states of the world' (see also Savage 1954, p. 13 and Eells 1982, p. 71-72).

of the world,⁷⁹ but that (unless the function is one-to-one) it needs not specify a *different* consequence for every state. In fact, it is a crucial assumption of the theory that every consequence can be realised in every state, so that there can be acts that work as “constant functions” by assigning the same consequence to every state.

Given these tools, Savage proceeds as follows to obtain theoretical representations of beliefs and desires. Firstly, he assumes that the agent in question has a preference ranking over the various acts she can choose (Savage 1954, p. 17). It is crucial to be aware of the implications of this assumption: Savage takes it for granted that the agent is fully capable of making up her mind about which actions she prefers. This could create some puzzlement, as it in a certain sense seems to beg the question the theory was meant to answer, namely how an agent is to decide amongst various possible actions. This is a point that I will return to below; for now, it is best to simply accept it.

Secondly, Savage lays down a number of postulates that any such preference ranking will have to satisfy in order to count as “rational”.⁸⁰ The most important ones amongst them for present purposes are P1, P2 and P4.

P1 states that the preferences amongst acts define a weak ordering, i.e. that it fulfils the requirements of dichotomy and transitivity (Savage 1954, p. 18). This seems entirely in line with what is to be expected from a theory of the kind that Savage is about to develop.⁸¹

⁷⁹ At least if the function is meant to be total, which seems reasonable given Savage’s general setup.

⁸⁰ To be exact, some of these “rationality postulates” – e.g. P6 and P7 – are really only technical requirements imposed to allow a neat solution to the problem; see also Suppes 2002, p. 252. I will not discuss this difference here however. How Savage’s notion of ‘rationality’ maps on to the distinction between logical and pragmatic rationality drawn in this thesis will be discussed further below.

⁸¹ He also hints at the possibility of merely setting up a partial ordering, but rejects further development of this idea as too complicated: Savage 1954, p. 21.

P2 is the most important and most controversial of the postulates. It is meant to codify what Savage calls the “Sure Thing Principle” (Savage 1954, p. 21-24). In brief, what the Sure Thing Principle states is that rational agents base their decisions between a set of acts only on those conditions where the acts differ and disregard all those where they are the same. Despite the seeming intuitiveness of the principle, it has been the object of frequent attacks. Some of these are presented in what follows below, so that I will postpone further discussion of them until then.

Next, P4 states if an agent prefers some prize f over some other prize f' (maybe because f involves more money), and if she prefers getting f if state B obtains and f' if not to getting f if state A obtains and f' if not, then this preference is not reversed if the prizes f and f' are exchanged for two different prizes g and g' which are similarly ranked by the agent⁸² (Savage 1954, p. 31). This principle is crucial, as it entails that the agent considers B more likely to occur than A. Given P5 (which postulates the guaranteed existence of prizes of the above sort) and P6 (which partitions the events in question appropriately), this is enough to establish the representation of beliefs by means of a probability function.

Finally, Savage completes the theory by adding one more structuring principle (P7), which allows him to show that the agent’s preferences over various gambles can – together with the previously elicited probability function – be captured in a utility function (Savage 1954, p. 73-81). The upshot is that an individual always acts by maximising her expected utility, where the expectation is calculated using her degrees of belief in the various possible states of the world.

⁸² I.e. g is preferred to g' .

This means that Savage's rationality postulates are meant to give the kind of justification for the expected utility maximisation that was lacking in Ramsey and de Finetti. Rational agents (i.e. those agents whose preferences satisfy Savage's postulates) will – by definition of their rationality – have beliefs that can be represented probabilistically, and will therefore be acting so as to maximise their expected utility.

3.2. Problems and Criticisms

In order to clarify what Savage's theory entails, where it can be considered successful (in terms of the topics at issue here), and where it faces severe difficulties, I present the most important objections that have been raised against it.⁸³

One may firstly wonder in what sense Savage's theory really is a "decision theory". The reason is that, at least on the face of it, it might seem reasonable to presume that a *decision* theory is meant to give the agent either normative guidelines for how to act in given situation, or to descriptively present the structure of her decisions.⁸⁴ In both cases, Savage's theory could be seen to fail spectacularly.

On the one hand, it seems descriptively false, as the large literature that was initiated by the seminal result of Allais 1953 appears to show.⁸⁵ On the other, its normative status is also dubious since (as was made clear above) the whole theory is *based on* the agent's preference ranking over acts. That is, a rational (in the sense of the satisfaction of Savage's postulates) agent is assumed from the outset to be fully capable of making

⁸³ Once again, the focus is squarely on what is important in the context of the present inquiry, which naturally means leaving out a host of otherwise interesting issues.

⁸⁴ An assessment that Savage seems to agree with: see Savage 1954, p. 19-21.

⁸⁵ This example is also discussed by Savage – see Savage 1954, p. 101-103; it is at this point not so relevant to assess to what extent his reply is successful, or what the 'paradox' actually shows. The most well known modern followers of Allais are Kahnemann and Tversky; see for example Kahnemann & Tversky 1972.

decisions about how to act. She does not need the theory to help her make up her mind about what acts she considers better and which worse in a given situation. This might cause one to wonder what the theory adds, given that it seems it can only derive its conclusion by assuming this very conclusion as one of the premises.

Here though is where the complexity of Savage's thinking has to be properly taken into account. In a certain sense, his theory could more tellingly be called an "endorsement theory". Its aim is *not* to tell the agent how to act, or to describe how she does act. Instead, it is meant to help her understand the *underlying structure* of her actions by explaining what her preferences entail about her beliefs and desires.

This, in turn, leaves her free to endorse or disavow these preferences. For example, she might now find her initial preferences to be mistaken in the light of the newfound clarity as to what they amount to and change them. On the other hand, she might find that even after careful consideration of the structure of her preferences, she is still committed to them. In this case, she can endorse her preferences, having gained considerable conviction that her decision stands on a solid foundation. In this way, she can come (maybe also by employing a method of reflective equilibrium) to have better knowledge of herself, her preferences and her actions.⁸⁶ Whilst falling short of satisfying the *prima facie* conditions of what a *decision* theory is meant to provide, this kind of knowledge can still be seen to be very useful.

It is also important to note that the question to what extent Savage has managed to elucidate the agent's decision making procedures is independent from his success in finding a representation of degrees of belief. Here, only this latter issue is of relevance; in

⁸⁶ There are some interesting issues here concerning preferences, preference-change and self-knowledge. However, since they are not directly relevant for the present thesis, I shall not discuss them further.

this way, the debate as to whether Savage's theory is appealing as an account of practical deliberation is important only to the extent to which it helps in understanding the nature of his method of representing beliefs.

The second worry focuses on what has come to be known as the problem of "state-dependent utilities" (see for example Bradley 2001, p. 284-285). As will become clear below, this is the central problem for Savage's theory, so that it is important to be clear about what it states.

It starts by noting that Savage's theory has – as an implicit assumption of the framework – the feature of being "state-independent". That is, in the derivation of his fundamental results, Savage has to take the desirability of the various consequences to be independent of the state in which they are realised (this was briefly mentioned above in the context of the constant function as a permissible 'act'). Agents are assumed to have a preference ranking over acts, and since the latter are merely functions whose values are the "consequences", this must mean that these preferences are based solely on the desirability of the consequences. In this way, the state-independence assumption translates all the way into the fundamental theorems (see for example Bradley 2001, p. 285 and Howson & Urbach 2004, p. 10).

Savage needs to appeal to state-independence in his definition of the qualitative probability measure (i.e. the "more probable than" relation) in order to obtain a *unique* probability function (see Savage 1954, p. 31-32). If it were dropped, then his system would only yield an equivalence class of probability measures, instead of a unique function (see also Schervish et al. 1990, p. 842). This would present a great difficulty for his method: as will become clear in the discussion of Jeffrey's account, non-unique

probability measures are highly problematic for a theory that is meant to represent beliefs by means of them. However, I shall postpone a more detailed treatment of this point until I come to it in the context of Jeffrey 1983, where it is more prominent and more easily discussed. For now, it is just important to be aware of the fact that the failure of state independence has highly undesirable consequences.

The objection surrounding the state-independence assumption then continues by noting the trivial but powerful observation that it seems to be simply false in most cases: being in a given state has profound influences on the desirability of a consequence. Going to the beach is presumably much more enjoyable on a hot, sunny day than on a cold, rainy one. Reading a book is more pleasurable in a quiet, bright environment than in a loud, dingy one, and so on. Examples of this sort abound. Any successful theory of these matters should be able to accommodate this trivial observation, however (as the objection continues), Savage's cannot. Thus, it is fundamentally inadequate. If it can only come to a unique representation of beliefs by making it, then it cannot ultimately be considered convincing (see for example Bradley 2001, p. 284-285).⁸⁷

By way of reply, there seem to be two options open. Firstly, Savage could introduce "conditional utility", i.e. make the desirability of a certain consequence explicitly dependent on the state it is realised in. It is however far from clear to what extent this is possible given the rest of Savage's framework: most of the attempts in the literature have

⁸⁷ Note also that Ramsey 1926 does not have this problem: his 'prospects' are not compatible with *any* given state; only ethically neutral propositions have this property. For more on this, see Bradley 2001, p. 285.

concluded that the chances for a coherent and compelling account that successfully amends Savage's treatment for state-dependence are dim.⁸⁸

Secondly, he could try to solve the problem by having a finer individuation of acts or consequences: instead of speaking of "having a bath", he would then speak of "having a bath on a hot day" and "having a bath on a cold day" as two fundamentally *different* consequences (see also Bradley 2001, p. 285).⁸⁹

The problem with this reply is that, once again, it is not at all clear that it can be successful. At the very least, it will have the downside of complicating the situation considerably (see also Jeffrey 1983, p. 20-22 for an illustration of this). More importantly, though, it is not clear that the added complications succeed in defending the theory against the above worry. For example, an ever more fine-grained description of the consequences might collide with Savages' requirement that any consequences can be realised in any state: it is unclear what it would mean for the consequence "having a bath on a hot day" to obtain when the state is "a cold day" (see also Schervish et al. 1990, p. 842).

On the whole, therefore, it seems that one has to conclude that state independence is very problematic for Savage's theory, as it is being relied on so heavily in its representation theorem, yet appears to be very counterintuitive.

There are two positive aspects connected to this problem however. On the one hand, the formal results of Savage's method are immune to these worries. The difficulties concerning state-independence are rooted in the conceptual foundations of the theory, not

⁸⁸ See for example Bradley 2001, p. 285, Howson & Urbach 2004, p. 10-11 and Schervish et al. 1990. Of course, this does not rule out that a successful amendment will be proposed in the future. It is merely meant to express the fact that, up to now, this has not happened.

⁸⁹ There is some suggestion that Savage himself preferred this last option: see Savage 1954, p. 25.

in its actual development. On the other, the main further problem for Savage's theory that is normally mentioned in the present context – his reliance on the 'Sure Thing Principle' – does not add much of significance to the above worry (see also Bradley 2001, p. 284-285). This can be shown as follows.

The problem with P2 (the formal equivalent of the "Sure Thing Principle") is that it is crucial for Savage's theory, but that it seems – contrary to his claims – far from being a rationality requirement. In fact, as much experimental work seems to have shown, this "rationality" postulate is being frequently violated in real life.⁹⁰ P2 might *appear* very convincing (after all, basing one's decision amongst acts on the states in which they differ, rather than on the ones where they are the same, surely does *seem* highly plausible), but that is misleading. For example, we might often consider acts only in combination with their alternatives, instead of assessing them on their own. Many theories have been proposed that try to explain – in this way or by other means – why the Sure Thing Principle is no rationality requirement.⁹¹

However, given the above considerations concerning state-independence, this problem seems far less damaging than it might at first appear. Firstly, Savage could reply (in a classic move) that one needs to distinguish between normative and positive decision theories. That is, Savage could simply claim that his theory is about how people *ought* to behave, instead of describing how they *actually do* behave.⁹²

However convincing such a reply may be found, there is a second and more important (in the present context) answer. This says that the alleged failing of the Sure Thing

⁹⁰ The 'locus classics' in this area once again is Allais 1953.

⁹¹ Often, this is done in the context of the 'independence axiom' in von Neumann & Morgenstern 1944 – see for example McClennen 2001.

⁹² In fact, Savage 1954, p. 19-21 does seem to suggest something of this sort.

Principle can really be put down to a violation of state independence, rather than the principle per se. This can be seen most easily as follows: an apparent failure of the Sure Thing Principle implies that comparisons between acts cannot be based solely on those states in which they differ, but must also take into consideration the states in which they have the same consequences. The most obvious reason for why that might be the case is that the various consequences are under-described and it is really that case that the same superficial consequences (“gaining \$100”, say) might mean very different things to the agent in different states of the world (for example if one starts out at \$0 and gains directly the \$100, or if one is first given \$200 and then loses \$100 – see also Joyce 1999, p. 101-103).

If that is true, then Savage could return to his previous reply to the objection concerning the state-independence assumption that if the consequences and acts were to be specified in enough detail, the alleged paradoxes might disappear (see for example Joyce 1999, p. 102 and Eells 1982, p. 76-77). Of course, this reply runs into exactly the same problems now as it did in respect to the violation state independence (e.g. that this might not actually be always possible). At least, though, what this shows is that the problems concerning P2 are not actually introducing any further difficulties here, but really reduce to those mentioned previously in the context of state independence.

In this way, the Sure Thing Principle remains an innocuous postulate perfectly on a par with the other – in some respects *technical* – assumptions Savage makes, *if one grants the truth of state independence*. It is this latter assumption then that turns out to be the main stepping-stone for the theory, and not P2 or any of the other postulates.

One final word is in order here about the notion of “rationality” that Savage employs, and how it fits to the two senses of the term distinguished in chapter I. Given the complex structure of Savage’s ‘endorsement theory’, it turns out that his notion of rationality comprises both pragmatic and logical elements. It is logical in the sense that the postulates of rationality are probably not satisfied by most actual agents, but that they should be seen to order the preference structure of an ideal reasoner. In this sense, Savage’s theory is really concerned mainly with the *logic* of decision making, and not so much with its pragmatics.⁹³

That said, however, there is an element of pragmatic rationality in Savage’s theory nonetheless. On the one hand, the theory is framed in terms that suggest a focus on the world of actual, non-ideal agents (events, acts, consequences etc.).⁹⁴ On the other, it helps one to learn about one’s preferences and their implications – in fact, it might help one find out more about what kind of person one is. This is pragmatically rational, as it is beneficial to know more about oneself (and, as made clear in chapter V, it might also save on the costs that come with the decisions one makes).

3.3. Conclusion

I have tried to show that Savage’s “endorsement theory” is a powerful and interesting theory that manages to provide the necessary normative justification (in the form of rationality postulates) for the development of unique theoretical representations of beliefs and desires.

⁹³ See for example Savage 1954, p. 7-8, p. 20 & p. 102-103.

⁹⁴ Note that I clearly do not want to suggest that events etc. are non-logical terms – only that logicians usually prefer to talk in terms of the more abstract propositions, sentences or sets.

Its main difficulty comes in the form of the assumption of state independence (one of its crucial parts), which though also spills over into problems surrounding its rationality postulates (particularly P2). Whether the theory can be successfully defended against these problems is still an open question, but at the very least it is clear that it represents a compelling step forward in the development of convincing foundations for a probabilistic representation of beliefs. In an attempt to push this development one step further, Jeffrey takes some of Savage's ideas as his starting point, but then departs significantly from Savage's framework to present an original take on these issues.

4. Jeffrey's "The Logic of Decision (2nd Edition)"

Jeffrey 1983's treatment set the tone for most of the modern philosophical work concerning the representation (and alteration) of beliefs; despite being similar in structure to some of the ideas in Savage 1954, it goes much beyond the latter in terms of the conceptual rigour of its foundations. This clearly implies that its way of representing beliefs needs to be studied carefully. In order to do this, I firstly present his account, and then contrast it with some problems that can and have been raised concerning it.

Before I begin discussing his approach, a brief further word is in order here. The one aspect of Jeffrey's theory that I pay relatively little attention to here is his own account of belief change. This is not because it is not a "good" account (on the contrary), but merely because it is not greatly relevant in the present context. The discussion here is limited to the *representation* of beliefs, and is not concerned with their alteration just yet. For this reason, I do not present the details of Jeffrey's account of belief revision until chapter V, where some of them are briefly contrasted with those of Bayesian Conditionalisation.

4.1. Jeffrey's "Logic of Decision"

Much of the terminology and basic set up of Savage 1954 is maintained in Jeffrey's framework, but there are some important differences in the details nonetheless. Jeffrey is also concerned with the combination of "acts" and "states" (which he calls "conditions") as to their possible "consequences" (see Jeffrey 1983, p. 1-2), but translates Savage's table from the previous section into *two* different matrices. On the one hand, there are the "probability matrices": these give the agent's degrees of belief (i.e. her cognitive

attitudes) in the occurrence of any condition, given her actions. On the other, Jeffrey introduces “desirability matrices”, which give the extent to which various combinations of acts and conditions are considered desirable by the agent (Jeffrey 1983, p. 2-4).⁹⁵

The important departure from Savage 1954 in this comes from the fact that the agent’s actions can make the occurrence of a state more or less likely – something that is ruled out in Savage 1954. As was made clear in the previous section, for Savage, state descriptions must not include any element that the agent has control over (as acts are functions from states to consequences, the acts cannot alter the likelihood of the state occurring without causing a breakdown of the functional setup). Jeffrey’s alternative treatment is not forced to draw this conclusion. Instead, he can see some actions as making some conditions more likely to occur: for example, bringing a football to a meeting in the park might make it more likely that a match will be organised. Jeffrey considers this a significant improvement over Savage’s much more complex treatment, which would have to deal with this situation with a more fine-grained individuation of states (see the previous section and Jeffrey 1983, p. 21-22)

It is also worthwhile to note at this point that the above characterisation of the desirability matrix should be taken with a pinch of salt: given Jeffrey’s “radical probabilism” (about which more is said below), probability-free judgements of desirability are not really possible. Any judgement of desirability must contain some element of a belief-based attitude on the part of the agent. I will however leave a further discussion of this point for later. What matters here is just that the agent somehow comes to a judgement as to the desirability of some consequence.

⁹⁵ This last aspect could also be made sense of in Savage 1954, as it merely represents one particular description of the world: see Savage 1954, p. 9.

The second move away from Savage's framework comes in Jeffrey's choice of terminology. Here, Jeffrey takes his cue from Ramsey and uses *propositions* where Savage's uses "events" and states. Since (as was mentioned earlier) *terminological* differences can be important guides to the underlying emphases of a theory, even if they do not correspond to real *logical* differences (see also Howson & Urbach 2004, p. 9), this shift in terminology points to Jeffrey's interest in giving a 'logic of decision' (viz. the title of his book). Crucially, though, Jeffrey also differs from Ramsey in that he 'unifies' the treatment of the cognitive and the 'pro'⁹⁶ attitudes in *only* talking about *propositions* (Jeffrey 1983, p. 59). Ramsey before him was concerned with both propositions and the somewhat mysterious "(conditional) prospects" (see above and Ramsey 1926, p. 77-78, Jeffrey 1983, p. 59).

Jeffrey (quite plausibly) sees this unification as clearly a beneficial element of his theory.⁹⁷ Two things relating to it need to be mentioned, however. Firstly, there are well-known difficulties with the ontological status of propositions (What are they exactly? Where are they exactly?). Whilst a deeper discussion of these difficulties is obviously beyond the scope of this inquiry, it seems clear that all of them also befall Jeffrey's theory.⁹⁸ This though is nothing *specific* about that theory, but merely a general fact of propositions. In this sense, there is no need to broach this subject in this context at all.

Secondly, it might be possible to reformulate Jeffrey's theory using other linguistic entities as the basic terms.⁹⁹ Jeffrey himself considers using *sentences* instead (Jeffrey p.

⁹⁶ This term is borrowed from Davidson 1963, p. 3-4.

⁹⁷ However, as will be argued below, this unification comes at a price.

⁹⁸ The same can of course be said about the *benefits* of talk about propositions.

⁹⁹ This would also make the difference between Jeffrey's and de Finetti's & Savage's vocabulary a starker *logical* contrast than before. The work of Davidson is also relevant in this context – see for example Davidson 1980.

64-65). This would have the benefits of bringing the theory closer to real life: different sentences can be used to express the same proposition, but it cannot be expected that *actual* agents assent to *all* the sentences that express the same proposition (Jeffrey 1983, p. 68). We often make mistakes and are not perfect logical reasoners. In fact, there is as yet no clear logic that would help to identify those sentences that express the same proposition from those that do not (Jeffrey 1983, p. 68).¹⁰⁰

What is very important about this point is that it immediately makes clear that Jeffrey is concerned with *ideal* reasoners (rather than actual ones) and the *logical* form of rationality (rather than its pragmatic one). In principle, ideal agents can be assumed to recognise the same proposition expressed in different ways (though actual agents might not¹⁰¹). By making propositions (and not sentences or utterances) the fundamental ‘unit’ of his inquiry, Jeffrey has thus set the stage for a ‘logical’ rather than a ‘pragmatic’ theory (Jeffrey 1983, p. 69).

The next major element of his theory is the derivation of the theoretical representation of degrees of belief and desirability as probabilities and “desirabilities”. Here, Jeffrey again follows the general structure of the previous attempts of Ramsey and Savage, though he differs in the details.

¹⁰⁰ This can be more precisely formulated as follows: there is no straightforward procedure that can be used to extricate a *proposition* from a *sentence*. For example, the sentence “Socrates is mortal and Plato is mortal” could be seen to express the proposition *that Socrates is mortal and Plato is mortal*, but it could also be seen to express the complex proposition *that Socrates is mortal and that Plato is mortal* (i.e. to be of the form A&B). There is no clear cut way of saying which is the “right” one, so that it is hard to check if the above sentence is equivalent to “Plato is mortal and Socrates is mortal”. Of course, once a correct “translation” into propositions is agreed on, mathematical logic can straightforwardly be used to provide effective procedures for tests of logical equivalence.

¹⁰¹ Note that this is a different worry from the one expressed in the previous footnote. There, it was shown that not even ideal reasoners might be able to extract the proposition expressed by a sentence from that sentence. Here, the worry is merely about the potential inability of actual agents to discern the same proposition if differently expressed (if one wishes, this is the reverse problem of that of the previous footnote).

Similarly to Savage, he starts by assuming that the agent has a preference ranking over a deductively closed set of the various propositions at stake (i.e. the Boolean algebra spanned by the initial propositions) – for example, she might prefer the truth of the proposition that they will be playing football in the park to that of the proposition that they will be watching TV at home. Amongst these propositions will be the tautology ($A \vee \neg A$) and the contradiction ($A \& \neg A$) (Jeffrey 1983, p. 76).¹⁰²

Given certain constraints on this preference ranking, two functions can be found that capture the probability and desirability dimensions inherent in this ranking (Jeffrey 1983, p. 148-149). As in the case of Savage, these constraints will consist partly of “rationality” axioms (like Jeffrey’s axiom 1: Jeffrey 1983, p. 145) and partly of technical requirements (like axiom 4: Jeffrey 1983, p. 147).

A crucial feature of the desirability function is the following “desirability axiom” (Jeffrey 1983, p. 80-81): the desirability of a disjunction of mutually exclusive propositions is the weighted average of the desirabilities of the individual propositions, with the weights being given by the ratio of the probability of the individual proposition being true to the probability of the whole disjunction. Formally, this can be expressed thus:

$$\text{des}(H \vee E) = \frac{\text{prob}(H)}{\text{prob}(H) + \text{prob}(E)} \text{des}(H) + \frac{\text{prob}(E)}{\text{prob}(H) + \text{prob}(E)} \text{des}(E)$$

¹⁰² For the existence theorem, the contradiction is removed from the field of propositions (it is left out of the desirability ranking): Jeffrey 1983, p. 146-148.

Moreover, some easy algebraic manipulations of this axiom show that in Jeffrey's theory, the desirability of *any* proposition is given by a weighted average of the desirability of the proposition being true in any relevant condition, where the weights are given by the probability of the condition obtaining conditional on the proposition being true.

To see this, consider some proposition H, and an exhaustive partition of propositions E_1 to E_n , with $(E_i \& H)$ and $(E_j \& H)$ being mutually exclusive for every i and j . Then H is identical to the total disjunction of all $(E_i \& H)$ (i.e. $H = (E_1 \& H) \vee (E_2 \& H) \vee \dots \vee (E_n \& H)$). Noting further that $\text{prob}(A \vee B) = \text{prob}(A) + \text{prob}(B)$ for two mutually exclusive propositions, the above desirability equation can be rewritten here as follows¹⁰³:

$$\text{des}(H) = \sum_{i=1}^n \frac{\text{prob}(E_i \& H)}{\text{prob}(H)} \text{des}(E_i \& H)$$

As $\frac{\text{prob}(A \& B)}{\text{prob}(B)} = \text{prob}(A/B)$, this yields the following equation:

$$\text{des}(H) = \sum_{i=1}^n \text{prob}(E_i/H) \text{des}(H \& E_i)$$

which expresses the above-stated desirability result formally. This axiom – together with the usual probability axioms - is then used to show that an individual acts so as to maximise her expected desirability over the set of propositions in question (Jeffrey 1983,

¹⁰³ This derivation follows Eells 1982, p. 79-80

p. 1).¹⁰⁴ Several features of this derivation of the probability and desirability functions (and their properties) are worth stressing in more detail.

Firstly, it is important to note the extent to which the above statement of the desirability of some proposition is related to Jeffrey's general philosophical outlook of "radical probabilism" (see for example Jeffrey 1991, p. 3). In general, "radical probabilism" refers to the view that probabilistic judgements are primitive. That is, it is not the case that probabilities at bottom have to be founded on *certain* pieces of evidence; for a radical probabilist, uncertainty starts right at the structural base of a judgement (Jeffrey 1991, p. 11).^{105, 106}

This outlook is also clearly visible in the nature of the above formulation: it shows that probability and desirability are conceptually and functionally related so that they cannot be separated. All desirabilities contain probabilities: given the above radical probabilism, there are no basic, belief-free value judgements.

Secondly, it is important to note that the representations of *both* desirabilities and probabilities are non-unique. That is, they merely form an infinitely large equivalence class of functions, instead of providing one unique such function. This is a feature of the theory that greatly distinguishes it both from Savage's (assuming state independence holds) and Ramsey's theory: in the two latter theories, only the utility functions are non-unique (i.e. allow positive linear transformations).¹⁰⁷ In Jeffrey's theory, however, *both*

¹⁰⁴ To be more precise: Jeffrey's theory implies that a rational agent maximises her *conditional* expected desirability (where the element of conditionality comes in because of the above manipulations yielding conditional probabilities in the statement of the axiom); see Eells 1982, p. 80.

¹⁰⁵ The second aspect of radical probabilism concerns the alteration of beliefs: given the above view of probabilistic judgments, there is clearly no requirement that one alter one's beliefs only upon learning a new piece of evidence with certainty; see Jeffrey 1991, p. 6. I will return to this point briefly below in chapters IV and V.

¹⁰⁶ For an entirely different justification of this kind of probabilism, see Joyce 1998.

¹⁰⁷ See also Eells 1982, p. 78.

desirabilities and probabilities admit of transformations (given the satisfaction of three conditions specified in Bolker's equivalence theorem: see Jeffrey, 1983, p. 97 and further below). Since these transformations will be discussed in more detail below (in particular with a view to possible standardisations of some of the above parameters), I shall not say much more about them here.

Thirdly, just as in Savage, the conceptual priority of the existence of the preference ranking means that the theory should be seen as an "endorsement theory". The agent is assumed to be able to make decisions concerning the desirability and probability of the truth of various propositions from the outset; Jeffrey's theory is there to elucidate the underlying structural dimensions of this ranking, so that the agent is then able to endorse or renounce it.

On the whole therefore, Jeffrey presents a philosophically rigorous attempt at combining certain aspects of Ramsey's and Savage's methods into a coherent and cogent whole. Its most important features lie in its being built on solid foundations of a Boolean propositional algebra, and in the radically probabilist framework in which it is set.

4.2. Criticism and Discussion

In order to clarify some key aspects of Jeffrey's theory, it is useful to discuss some of the classic objections in the literature that have been raised against it. First and foremost amongst those stands the non-uniqueness of the belief representations that his theory yields. As was made clear earlier, Jeffrey's theory admits certain kinds of transformations of the probability and desirability functions. Taking a closer look at these will make clear exactly where the problem lies concerning the non-uniqueness of Jeffrey's belief

representation. The allowed transformations are of the following form (where prob' and des' are the transformed functions from the initial functions prob and des; see Jeffrey 1983, p. 97):

$$\text{prob}'(H) = \text{prob}(H) (c \text{ des}(H) + d)$$

$$\text{des}'(H) = \frac{a \cdot \text{des}(H) + b}{c \cdot \text{des}(H) + d}$$

At this point, it is also useful to write out the three conditions of the equivalence theorem mentioned earlier that have to be satisfied (see Jeffrey 1983, p. 97):

(a) $ad - bc > 0$

(b) $c \text{ des } X + d > 0$ (for all X)

(c) $c \text{ des } T + d = 1$ (where T stands for a tautology)

The non-uniqueness of the probability function can be brought out most easily if it is noted that many of the above parameters can be normalised (see also Jeffrey 1983, p. 99-100). For example, one could set

$$\text{des } T = 0.$$

Then by condition (c), it immediately follows that $d = 1$.

Further, from the desirability transformation, we know we must have

$$0 = \text{des}' T = b/d = b.$$

If some proposition A is moreover chosen as a “unit”, so that

$$\text{des } A = 1,$$

then by the same reasoning as above it follows that $1 = \text{des}' A = a / (c+1)$, which implies that $c + 1 = a$.

On the whole, this means that

$$\text{des}' X = \frac{(c + 1) \cdot \text{des}(X)}{c \cdot \text{des}(H) + 1}$$

and

$$\text{prob}' (X) = \text{prob} (X) (c \text{ des} (H) + 1)$$

What this means is that even after choosing a zero point and fixing a unit, the function is still non-unique when it comes to assignments of probabilities. That is, the same set of preferences can be theoretically captured in an infinity of ways: the theoretical representations of the desire and belief components underlying them merely define equivalence classes, not unique functions.¹⁰⁸

The problem with this is that it entails that the same preference ordering can lead to probability functions that disagree about the ordering of some propositions in the ranking. Moreover, this in turn implies that some agent’s degrees of belief concerning some sets of propositions might be thus that they cannot be captured by *any* probability function

¹⁰⁸ The above is true unless there are no finite bounds on the desirabilities (which is somewhat implausible: at some point, our capabilities of distinguishing good or bad states of affairs seem to give way; moreover, unbounded preferences run into problems connected with the St. Petersburg Paradox; see also Joyce 1999, p. 137), then c is driven to 0 due to condition (b) above: see Jeffrey 1983, p. 100-102.

(see for example Joyce 1999, p. 136). These are fairly disastrous upshots in the present context: the whole point of setting up Jeffrey's complex framework was to find some way of theoretically representing beliefs. If the theory fails in this – either because no representation exists or because it leads to inconsistencies – then the theory, at least from a logically rational point of view, cannot be considered adequate.

To see this more clearly, notice how awkward the non-uniqueness of belief representations is: if it is truly the case that probabilities can be seen as representing degrees of belief, then it would seem to suggest that different probabilities represent different degrees of belief. This, however, does not hold for non-unique probability functions. It is questionable, therefore, whether probabilities really represent or measure degrees of belief at all in these cases.

This can be seen easily by comparing the situation of the non-unique probability function with that of the equally non-unique desirability function: this last non-uniqueness is a feature of almost all theories proposed in this context – yet is not seen as a problem at all. The reason for this is that this latter function merely represents a desirability ranking such as “I desire X more than Y”, where it is irrelevant whether X and Y are given values of 10 and 5 or 1004 and 23. All that matters is that X is given a higher value than Y.¹⁰⁹ However, the situation is very different for degrees of belief: there, it is normally assumed that they do come on a fixed scale (at least in principle): I do not just believe X more than Y, but I do so in a definite ratio. Jeffrey's non-uniqueness result does away with this natural and intuitive assumption.

¹⁰⁹ This non-uniqueness becomes a problem only for the interpersonal case, which I however shall not consider here.

Three replies can be given to this problem. On the one hand, Jeffrey could bite the bullet and say that these are indeed the consequences of his theory, but that these consequences are not as disastrous as they are made out to be (for a reply along these lines, see Jeffrey 1983, p. 161-162). For example, he could claim that given his radical probabilism, it is just to be expected that degrees of belief might permit of different orderings of some propositions. If these judgements are basic, then there is nothing to choose between them: people might simply disagree. Just as it is useless to argue about differences in taste, it is useless to argue about differences in basic probabilistic judgements. The extent to which this reply is convincing is still open to debate, however (also because it would seem to imply that there really is no fact of the matter what an agent's degrees of belief are, which is a very strong claim).

The second answer he could give is to place further constraints on the assignment of degrees of belief: it is relatively well known how to do this in order to achieve the desired uniqueness of the probability function (see for example Joyce 1999, p. 137). The problem here is that this would essentially alter the whole approach of eliciting a representation of beliefs from a preference ranking. If there are "rational degrees of belief", then Jeffrey has in effect returned to something closer to the Logical Theories of chapter II than the more subjectivist attempts of this chapter.

A third way of obtaining the desired uniqueness is to introduce preferences over the truth of special kinds of conditionals, which give enough structure to the theory to provide a unique probability function (see Bradley 1998; for a brief criticism, see Joyce 1999, p. 138 note 15). It also turns out that these conditionals satisfy "Adam's thesis" about conditional propositions: the probability of such a proposition is simply the

probability of the consequent conditional on the antecedent. The details of this proposal are not so relevant for the present thesis; what matters is that this way of trying to solve the problem of non-uniqueness also comes at a cost. The first such cost is that the abovementioned Adam's thesis has been severely criticised (see for example Joyce 1999, p. 191). The second is that the way an agent's preferences over these conditionals have to be structured can be called into question (see also Joyce 1999, p. 138 note 15).¹¹⁰

On the whole, it can be concluded that Jeffrey's theory faces some severe *logical* difficulties, with it being unclear whether Jeffrey can come to compelling answers to them. However, this does not detract from the fact that, on a pragmatic level, it remains a fairly convincing account: just like Savage 1954, Jeffrey can at least claim that in a number of circumstances, his endorsement theory will tend to yield beneficial results to the agent, by helping her get to know the structure of her preferences.¹¹¹

The second problem for the theory concerns its notion of an "act". The problem here is that it is far from clear in what sense an individual's (intentional) *actions* can really be captured in it: since the theory is completely framed in terms of propositions, it must somehow be possible to characterise an act in terms of preferences over the truth of these propositions. Jeffrey himself maintains that the notion of a preference over a set of propositions admits of two different interpretations: on the one hand, it can be seen to be about "news reports". This is to mean that, on this interpretation, an agent's preference of proposition A over proposition B is to be understood as her preferring to learn that A obtains over her learning that B obtains (Jeffrey 1983, p. 82-83).

¹¹⁰ For a good overview over these and related issues, see Eells & Skyrms 1994.

¹¹¹ See also the end of this chapter.

On the other hand, where it is in the agent's power to make both A and B true, a preference of A over B can also be understood as a preference over the acts that these propositions characterise (Jeffrey 1983, p. 83-85). For example, if A is the proposition that I go for a bike-ride tomorrow, and B the proposition that I take a walk tomorrow, then my preference of A over B can roughly be characterised as my rather going for a bike-ride than going for a walk tomorrow (Jeffrey 1983, p. 83).

However, this way of accommodating intentional action within Jeffrey's theory has been called into question. The difficulty here comes out clearly if it is noted what use these preference rankings are put to in Jeffrey's theory: namely, to derive representations of degrees of belief and desirability. This means that an agent's degree of belief in some action of hers is less than certainty: if a preference ranking over propositions can be (partially) captured in terms of a probability function, then some of these propositions will have a higher probability value assigned to them than others. For a set of propositions that characterise acts, this would mean that the agent has different degrees of belief in the acts happening – which seems false. If I decide to do something (say, to raise my arm) then I am *certain* that it will happen; if I decide not to, then I am certain that it will not. The upshot of this is that, since acts only permit of probability 1 or 0 (whereas propositions in Jeffrey's theory can take any value), acts cannot be accounted for as easily in Jeffrey's theory as he claims.

One answer that seemingly provides a way out of this difficulty is the following (Jeffrey 1983, p. 83 hints at something similar). It is simply *false* that an agent can be certain an act of hers actually “succeeds” or “happens”. There is always the possibility that something occurs that thwarts the completion of the intended action: for example, I

might hit my hand against a table, thus preventing my arm from going up. The only thing that the agent can truly be certain about is her *trying* to act in a certain way: I can surely *try* to raise my hand – but there is always the possibility that I may fail in the attempt (though that possibility might be very small). This explains why an agent can have degrees of belief different from 1 and 0 in the truth of some act-proposition.

The problem with this reply is that it seems to run into the same problem again, only this time in terms of *trying*. For example, consider the propositions that I will try to take a bike-ride tomorrow (A') and that I will try to take a walk tomorrow (B'). Then it seems that Jeffrey's theory might entail that I have a degree of belief of less than 1 in my trying to take a walk. That though seems fairly strange: how can I be unsure as to whether my *trying* to do something might succeed? It would seem to entail that one can only try to try to do something (a kind of second-order trying), which appears ludicrous.

It is not very clear how Jeffrey could reply to this renewed onslaught of the objection. One possibility is for him to bite the bullet; in fact, he could do this in two ways. Firstly, he could claim that given this objection, his theory is really only to be understood as being based on news-items. Given its abovementioned construal as an endorsement theory (rather than a decision theory), this might not be too damaging. Secondly, he could claim that, as unintuitive as it might appear, there is really no certainty in anything, even "trying". This in turn might fit well to his general philosophy of radical probabilism above. In what sense these replies are convincing is once again a matter of debate; however, since the present thesis is in an inquiry into the representation of beliefs, these issues do not matter greatly here. These are problems about the cogency of Jeffrey's

theory in terms of a decision theory, not as a method for representing beliefs. I shall therefore leave this issue aside here.

The final difficulty that needs to be briefly mentioned here is the lack of causal connections in Jeffrey's theory. Jeffrey believes that there is no need to account for causal notions *within* one's decision theory. This is not to say that causal notions are not important, but merely that they need not be taken for primitive elements in one's theory (see for example Jeffrey 1983, p. 157 and Eells 1982, p. 95). In order to illustrate this point briefly, consider the following example (adapted from Eells 1982, p. 89-92).

Assume that there is a high positive correlation between chocolate consumption and divorce, but that this is due to there being a common cause of both: frequent fights amongst the spouses. Consider the case of Susie, who feels like having chocolate. Knowing about the above correlation and being a keen follower of Jeffrey's theory, she might reason as follows.¹¹² Since the probability of a divorce given her eating chocolate is high and given that the desirability of a divorce is low, she decides to follow Jeffrey's expected desirability maximisation, and concludes that it is best not to eat the chocolate.

The problem with the above reasoning is that it seems simply false: since Susie both believes that chocolate consumption has no influence on the success of her marriage and would like to eat some, it seems foolish not to. That though is exactly the course of action advised by Jeffrey's theory. In order to remedy this defect, causal decision theorists advise a return to something similar to Savage 1954, however with causal notions built right into the theory (see also Eells 1982, p. 95).¹¹³

¹¹² The next sentence merely rephrases in words what would be given by the equation in terms of conditional probabilities above.

¹¹³ One of the classical formulations of this view is Lewis 1981; a good overview over these theories is given by Joyce 1999.

Whilst a deeper discussion of these causal decision theories is beyond the scope of this thesis, a few remarks are in order here. Firstly, one of the problems with a return to the ideas of Savage 1954 is that many of the problems of the previous sections return as well. Secondly, it is not fully clear that an amended version of Jeffrey 1983 could not deal with the above situations more adequately. For example, Jeffrey could claim that the above cases based on common causes do not represent decision situations at all – if these common causes are really as strong as the examples make them out to be, then there is no need for Susie to deliberate about what she is to *do*, as her actions are predetermined anyway. That is, they are not really actions at all.¹¹⁴ Here as before, the jury is still out as to how compelling objection and replies really are here. Once again, though, this is a problem that affects Jeffrey 1983 as a decision theory, and not as a means of representing beliefs, so that I will leave this issue aside here.

Summing up: as the title of Jeffrey 1983 makes clear, Jeffrey's theory shows how the preferences of logically rational ideal reasoners can be used to come to a representation of their beliefs and desires (the fact that he is concerned with ideal reasoners comes out clearly for example in his reliance on propositions in framing his theory). As was the case in Ramsey (in many ways the precursor of Jeffrey's theory – see for example Jeffrey 1983, chapter 3), his take on these issues is equally about the structure of preferences and propositions, and not so much about the behaviour of actual agents. This of course does not mean that his theory cannot also be useful for real-life agents – it just means that if it is useful in this way, then that is because of its *pragmatic* properties, not its *logical* ones. More about this will be said at the end of this chapter.

¹¹⁴ For yet another response, see Eells 1982, chapter 6.

4.3. Conclusion

I have tried to show that Jeffrey's theory deserves to be one of the most popular and well-known philosophical methods for the representation of desires and beliefs in a coherent theory. Its main problems lie in the non-uniqueness of these representations, and the fact that in its strong reliance on the logic of propositions for ideal reasoners, it struggles to account for intentional actions and deliberations of actual agents. This, however, has not stopped it from setting the tone for much of the work that is currently done in the field; in particular, it can in some ways be seen as the starting point of Howson & Urbach's more recent treatment. As argued in the next section, this takes a number of Jeffrey's conclusions yet another step forward with a view to exploring the similarities between deductive logic and the probabilistic representation of beliefs.

5. Howson & Urbach's "Scientific Reasoning – The Bayesian Approach"

Howson & Urbach's treatment is one of the most well known *current* approaches to the issues of this chapter. Apart from this, there are two further reasons why it is an instructive choice as a representative of the modern work in this area:¹¹⁵ on the one hand, it continues and expands the paths that had been taken before, and on the other, it shows the fruitfulness of possible applications of it. This combination of traditional ideas with current problems presents a concise summary of the present state of the debate.

Furthermore, it is also worthwhile to take a brief look at the way Howson & Urbach seek to apply the representation of beliefs to current statistical practice in science. This not only introduces a debate that is interesting in its own right (though largely perpendicular to the issues at stake in the present thesis), but it also makes clear how important and wide-ranging the consequences are of a thoroughgoing commitment to seeing probabilities as the representations of degrees of belief.

Howson & Urbach's method of belief-representation is presented in the first part of this section, and then discussed in the second. I finally conclude with an overview of what Howson and Urbach have added to the state of debate in the third and last section.

5.1. *Howson & Urbach on the Representation of Beliefs*

The first point to note in respect to Howson & Urbach's take on the probabilistic representation of beliefs is that they leave behind the general framework of a causal

¹¹⁵ Other works to be mentioned here are Skyrms 1986, Earman 1992, and Joyce 1999. It should also be noted that by confining myself here to the work of Howson & Urbach, I am not making an implicit statement about the importance (or lack thereof) of these other treatments, to which I shall continue to make frequent references.

theory of intentional action that was so prevalent before. That is, instead of trying to derive an agent's degrees of belief from their actions or preferences, they want to tackle the problem more directly and base the representation of beliefs on the agent's judgements concerning the "fair odds" of a bet (Howson & Urbach 1993, p. 76-77). More about this will be said below, but for now it is important to keep in mind this shift in the structure of the account.

Secondly, implicit in this reliance on "fair odds" is that they do follow de Finetti 1937 in taking the paradigmatic situation for the elicitation of beliefs to be *bets*: just like him, they use the betting quotient¹¹⁶ of a bet as a representation of an agent's degree of belief, and then show that betting quotients are probabilities (Howson & Urbach 1993, p. 89).

There are two main points of departure from de Finetti's previous results however: firstly, they do not frame their account in terms of the notion of "events", but instead concentrate directly on a logic of propositions in the vein of Jeffrey 1983 (see Howson & Urbach 2004, p. 19). This difference in terminology once again betrays a deeper difference in the emphasis of the theory, as will be made clear below.

The second point of departure from de Finetti comes in a return to Ramsey 1926 and the reliance on *hypothetical bets* only (Howson & Urbach 1993, p. 77). That is, unlike the former, they envision asking the agent in question merely what they *would* consider a *fair* bet, given that the truth of the various propositions will be revealed at the end of the bet, and that there are no actual wins and losses involved (Howson & Urbach, p. 90). This allows them to measure degrees of belief as the betting quotients for hypothetical bets an

¹¹⁶ Note that Howson & Urbach prefer to talk about odds rather than about betting quotients; however, these are merely two sides of the same coin: the former is related to the latter by $q = \frac{k}{1+k}$, where q is the betting quotient and k are the odds of the bet (see Howson & Urbach 1993, p. 76).

individual considers fair. This methodology is at the heart of their approach to all of the issues surrounding the representation and alteration of beliefs, so it is important to be clear about several of its details.

First amongst them stands the clear rejection of behaviourism (Howson & Urbach 1993, p. 77): like Ramsey 1926, they see no difficulties in basing their approach on an agent's verbal expressions of their hypothetical choices, but unlike him, this does not arise out of adherence to some form or other of a behaviourist philosophy.¹¹⁷ Instead, this is because they are interested in drawing out the consequences in terms of the coherence of a set of judgements of an agent (i.e. what the fair odds of a bet are); for that, they naturally have to rely on the agent's expression of these judgements, whereas her behaviour is secondary.¹¹⁸ The contrast between Ramsey and Howson & Urbach could therefore be summarised as follows: Ramsey is mainly concerned with *behaviour*, but is tolerant about that being merely hypothetical, whereas Howson & Urbach are mainly concerned with *hypothetical choices*, but are tolerant about these involving behaviour.

Of course, given this reliance on hypothetical choices, they have to face the same difficulties that Ramsey did concerning an agent's introspection about her *strengths* of beliefs and her hypothetical behaviour. However, since these have been discussed already in connection to Ramsey's method, I will not deal with them any further here. It should however be kept in mind that everything said about it in the context of Ramsey 1926 is applicable here also.

¹¹⁷ One of the reasons for this is probably that – as argued in section 1 above – behaviourism as a philosophical position is considered somewhat outdated: see also Gillies 2000, p. 138-141.

¹¹⁸ This is not to deny that these judgments might express themselves in some sort of outwardly observable action. However, whether they do or not is simply not relevant for their method of belief representation: see Howson & Urbach 1993, p. 77.

The second point to note in Howson & Urbach's methodology concerns the kind of epistemological idealisations involved. Here, it is best to illustrate with an example:¹¹⁹ consider some proposition L that describes some universal law-like statement (for example, that all ravens are black). In any actual bet on the truth of L, no pragmatically rational agent would be willing to have a betting quotient of anything other than 0, simply because she would never win if L were true¹²⁰ and definitely lose if L were false. Therefore, if one were to take actual betting quotients as a measure of an agent's degrees of belief, one would have to conclude that her degree of belief in L is 0.¹²¹ This, however, need not be the case: the 0 betting quotient is an artefact of using actual bets and does not represent accurately the individual's cognitive attitudes.

This situation is considerably different for Howson & Urbach's methodology of hypothetical bets: there, the individual might still name a betting quotient other than 0 as a fair bet on the truth of L, despite this being *pragmatically* irrational: as Howson & Urbach are concerned with the logical rationality of an ideal agent, this is irrelevant.¹²² This agent merely has to name a betting quotient she would consider fair, on *the assumption that the truth or falsity of L will be revealed once she has done so* (Howson & Urbach 1993, p. 90).

To see clearly why they appeal first and foremost to ideal reasoners (as required for arguments concerning logical rationality – see chapter I), note that this stance involves

¹¹⁹ This was originally raised as an objection to a Bayesian philosophy of science by Popper 1999, appendices vii&viii; it is also discussed in this respect by Howson & Urbach 1993, p. 395.

¹²⁰ Assuming there is no known limit to the number of ravens in the universe. More generally, this is simply a consequence of the problem of induction: barring an appeal to the metaphysical principle that nature is uniform, no (finite) amount of observations is enough to establish the truth of a universal claim: see Hume 2000, book I, part II.

¹²¹ Using less of an operationalist terminology: a 0 betting quotient fails to *measure* accurately the agent's beliefs.

¹²² For more on the relevance of this distinction to this question see section 6.2. of this chapter.

considerable epistemological idealisations: it might be impossible for any human being (either currently or in principle) to reveal the truth of L. However (keeping in mind the distinction between logic and pragmatic rationality), what matters for Howson & Urbach is that this information could *in principle* be revealed – not whether it can *actually* be done. This situation is comparable to that in deductive propositional logic: there are effective procedures for checking whether some proposition P follows from a set *Q* of other propositions (Enderton 2001, p. 63). This however does not mean that it is also *practically* possible to test this: actual human agents will find the “effective procedures” specified to be inapplicable in real life.¹²³

The third and final crucial element of Howson & Urbach’s methodology is in their notion of ‘fairness’: they argue that it is historically and intuitively the case that a ‘fair bet’ is one that has a zero expected value, i.e. confers no advantage to either side (Howson & Urbach 1993, p. 84-85). Given this notion of fairness (which, once again, does not commit them to seeing an agent as always willing to take fair bets: see Howson & Urbach 1993, p. 77), their approach can be summarised as follows: they seek to inquire into the relations that have to hold between various betting quotients if they are to remain fair. In other words, they want to establish a ‘logic of fair betting quotients’, so as to show what an agent’s judgements concerning a set of fair odds commit her to (assuming she is logically rational) in terms of other suchlike judgements.

Given these philosophical foundations of their method, the formal aspects follow immediately. If an agent’s beliefs can be captured in her judgement concerning the fair odds of an hypothetical bet, then by transforming these odds into betting quotients, de

¹²³ This is true in all but the simplest cases when checking the validity of an argument: see Enderton 2001, p. 61.

Finetti's theorem from above can be applied unchanged to show that these betting quotients have to satisfy the axioms of the probability calculus in order to remain fair (Howson & Urbach 1993, p. 79-84). Altogether, this means that, once again, beliefs can be represented using probabilities.

Before I begin scrutinising this take on the representation of beliefs more closely, it is at this point very instructive to briefly introduce another contemporary debate in this area. This debate moves beyond the argument of this thesis by asking: if probabilities can truly be seen as the representations of beliefs, what does this imply for the scientific *use* of these probabilities?¹²⁴ Apart from showing the concrete consequences of the issues raised in this inquiry, asking this question is also important for drawing out once more the links between the representation (and alteration) of beliefs and contemporary science.¹²⁵

To see the consequences of seeing probabilities (at least sometimes) as representing degrees of belief, it is best to consider briefly the common statistical practice of significance testing as a contrast.¹²⁶ This methodology is predicated on seeing probabilities as objective features of the world, and *not* as coding beliefs. What it aims to do is to use the properties of the available data in such a way as to draw conclusions about the underlying ('objective') processes or populations from which the data was drawn. A crucial part of this methodology is that it seeks to avoid any subjective elements *in the procedure itself*.

¹²⁴ As was pointed out before, the present thesis is not so much concerned with the nature of probability as such – whether it is just subjective, or whether it sometimes also has to be construed in objective terms by employing frequencies or propensities. For an overview over these issues, see Gillies 2000.

¹²⁵ What follows below is not meant as a complete discussion of Bayesian statistics, but is intended just as an *illustration* of the practical consequences of the probabilistic representation of beliefs.

¹²⁶ For clear and concise treatments of these issues, see Newbold et al. 2003, chapter 9 and Dougherty 2002, chapter 3. Note also that the version of classical statistics here presented should not be expected to be entirely congruent with the versions of the founding fathers of the approach (Fischer, Neyman, Pearson). The aim is instead to reduce the theory to its barest essentials so as to avoid unnecessary complications, whilst however keeping it recognisably "classical".

Despite being extremely widespread, this practice therefore is fundamentally at odds with seeing probabilities as representing degrees of belief. If this latter position is taken seriously, then the crucial element missing in all of the above (apart from a subjective interpretation of the notion of ‘probability’) is the statistician’s “prior probabilities”, i.e. her initial degrees of belief (Howson & Urbach 1993, p. 353).¹²⁷ If these were taken into account, a whole array of new possibilities could be opened up: on the one hand, by allowing proper application of Bayesian Conditionalisation as a way of expressing the effect of evidence on scientific theories (as will be shown in the next three chapters), and on the other, by avoiding some difficulties of the testing approach (see for example Howson & Urbach 1993, p. 243-244). It however does require a somewhat different statistical methodology, by introducing the scientists’ beliefs directly into the methodology. This though might not be a *vice* of the procedure, but a *virtue*: if science truly is about the rational revision of beliefs,¹²⁸ then these beliefs should be taken into consideration in the very methods that scientists use. Note also that one can – as many statisticians in fact do – fall back on further ‘objective’ constraints on these beliefs to make this procedure less subjective in outlook (Objective Bayesianism is again one of the most recent suchlike approaches: see Williamson 2005).

Thus, the benefit of changing statistical methodology in this way is not only that there is no “loss” involved (as all of the same things a classical statistician can do a “Bayesian” can do as well, due to modern computational powers – see Howson & Urbach 1993,

¹²⁷ Note that Howson & Urbach 1993 also collate an array of specific objections to the classical approach above: for example, the non-uniqueness of test-statistics, its notion of ‘rejection’ and its reliance on “stopping rules” (see Howson & Urbach 1993, chapters 8-13). These are not greatly relevant in the present context, however, so I shall not debate them here.

¹²⁸ This will also be picked up in chapter VI.

chapter 14), but also that it would then be based on a much sounder philosophical footing (Howson & Urbach 1993, p. 381).

On the whole, therefore, Howson & Urbach present a treatment of the representation of degrees of belief by probabilities that is based on a logic of fair betting quotients about hypothetical bets. They also show the fruitfulness and power of these ideas for many fields of probabilistic reasoning.

5.2. Problems and Replies

Just as before, a complete understanding of the approach requires also an overview over the main difficulties of the theory, and how they could be overcome.

The first problem focuses on an issue right at the heart of their methodology: given the above arguments, it might be conceded that Howson & Urbach have shown that *fair odds* have to satisfy the axioms of the probability calculus (and thus can be analysed in terms of their logical coherence), but that they have not really provided an argument for why fair odds should be seen as a measure of an agent's beliefs *in general*. It might be more plausible to claim that their system shows that one particular kind of belief (namely, about fair odds) can be represented probabilistically. However, this is very different from claiming that beliefs in general can be so represented. This difficulty in this comes from abandoning the causal theory of intentional action: because of *it*, de Finetti (say) could claim to have represented an individual's degrees of belief – if actions are caused by desires and beliefs, then these actions can (in certain circumstances) be used to come to a representation of the agent's beliefs. Howson & Urbach do not have this option open to

them, due to their refusal to rely on the causal theory of intentional action and any overt behaviour of the agent.

However, Howson & Urbach have a reply to this objection: an agent's assessment of the fair odds of a bet on the truth of some proposition clearly depends on that agent's *beliefs* about the truth of that proposition (Howson & Urbach 1993, p. 76). This then is the reason why the method is so cogent: as beliefs translate so readily into judgements of fair odds, and as the coherence of fair odds requires them to be probabilities, beliefs are straightforwardly representable as probabilities.

There is slightly more to this objection than the reply would let on however: the notion of "fairness" involved in the hypothetical bet might be far less amendable to translations into degrees of belief than it might at first appear. In a certain sense, this is the analogue problem to that of de Finetti's betting approach earlier: it seems that the notion of 'fairness' as equal advantage to both sides of a bet might be seen to be too specific to be generally applicable.¹²⁹ For example, some cultures might have a very different conception of a 'fair bet': maybe people in positions of authority should be given more (or less) weight in the two sides of the gamble in order for it to count as fair, or maybe 'fairness' in bets is tied to the stakes in question (for example in relation to the total wealth of the individuals involved in the bet) instead of the terms of the bet. If that is so, then the correspondence between degrees of belief in the truth of some proposition and the judgement about the fair odds of a hypothetical bet on the truth of that proposition breaks down. Now, the judgement of the 'fair odds' also contains the agent's beliefs

¹²⁹ Note also that in a certain sense, speaking of an "advantage" here might be considered out of step with the rest of Howson & Urbach's methodology: it seems that implicit in this notion is an appeal to the *utility* of the agents in question. This though would reintroduce the causal theory of intentional action into their account – and as made clear above, this is something that Howson & Urbach want to avoid: Howson & Urbach 1993, p. 77.

about the ‘fairness’ requirements of the situation, and her general cultural, social and personal situation. This might work as a disturbing cause (similarly to de Finetti’s system) to disrupt the representation of beliefs.

Once again though, Howson & Urbach might be able to meet this objection. Firstly, they could note that the historical evidence available points to seeing the fairness of a bet truly in terms of the advantage conferred to the various sides of the bet (see for example Hacking 1975, p. 92). Secondly, even if that were not the case (or if evidence were discovered later of a different conception of fairness in some circumstances), they could try to defend their method by being very explicit about the definition of “fairness” *they* use. For example, if some culture would take the initial wealth of the bettors into consideration when determining whether the odds of some bet are “fair”, Howson & Urbach might then either drop the label “fair” from *their* odds, or make it explicit what *they* mean by that term. It would seem possible for an agent with a different conception of the “fairness” of some odds to make the mental leap towards Howson & Urbach’s concept with that name; moreover, this might be enough to represent that agent’s beliefs by means of probabilities.¹³⁰

The second objection to Howson & Urbach’s account has to do with the hypothetical betting situation they envision. The problem here is that *asking someone what they would take to be fair odds* and *asking them to stake some actual money on these odds* are two fundamentally different situations. The difficulty for the former is that it might not have enough normative force to bind the agent into a framework that yields reliable measurements of her degrees of belief, due the lack of the ‘economic channelling’ that

¹³⁰ These two claims might be called into question. However, these are ultimately issues for an empirical investigation of these matters, so I shall not debate them further here.

actual bets provide. The problem is that there is no penalty whatsoever on the agent for naming odds that she does not “really” take to be fair (maybe because she did not deliberate about the situation sufficiently). This may mean that the fair odds that the agent names do not truly represent her beliefs at all, and are really just numbers quoted from the top of her head.¹³¹

By way of reply, Howson & Urbach could make two claims. Firstly, they could simply bite the bullet and claim that agents are fully capable and willing to make these judgements. Human beings (and especially the ideal agents they have in mind) are not simply conditioned dogs that have to be punished or rewarded for everything they do. Despite the absence of monetary losses, therefore, fair odds might provide the needed foundation on which to build a representation of beliefs.

Secondly, they might note that the subjects of their account are ideal reasoners, and their aim a ‘logic of beliefs’. It is once again instructive to compare this to the situation in standard deductive logic: in this field also, there are no direct ‘penalties’ for assigning truth values to propositions in a way that one does not think fully accurate. Deductive logic is about what further truth assignments an agent is committed to, given that he has already made a number of them. Equally here: Howson & Urbach want to show what further beliefs an agent is committed to, given that he holds certain other ones (Howson & Urbach 1993, p. 77 and Howson & Urbach 2004, p. 19). In this respect, it is irrelevant what beliefs the agent “really” holds: *that* is an issue of pragmatic rationality. What Howson & Urbach are interested in, however, is the question what follows for an agent’s

¹³¹ Notice the difference to Ramsey, who does not have this problem to this extent, as his hypothetical choices play a fundamentally different role in the representation of beliefs.

beliefs, given that she holds certain other beliefs; this in turn concerns that agent's *logical* rationality.

There is one last set of difficulties with Howson & Urbach's account that has to be discussed here. This one centres on the possibility of actually *naming* the fair odds for *any* situation, which seems to be a very strong assumption. That is, it seems quite plausible that there are many situations where it might be very difficult for some agent to name what her fair odds would be. For example, a biologist might be incapable of saying what she would consider a fair bet on the discovery of a gene that causes depression. Maybe there is not enough known about the situation to make any but the most wild guesses possible, or maybe the available evidence is just so contradictory that she feels unable to make any guess whatsoever.¹³² This makes this notion once again highly inadequate as a foundation for a compelling theory of the representation of beliefs: it is quite possible that, despite the difficult evidential situation, she still has beliefs about the matter. It is questionable however whether she can name the precise fair odds that are meant to code them.¹³³ Therefore, fair odds might fail as cogent representations of her beliefs.

However, it is once more crucial to keep apart logical and pragmatic rationality here. Clearly, the above might very well be true for actual reasoners: *they* might be unable to specify what the fair betting quotients are in every given situation. However, the system above is meant to apply to ideal reasoners: it is about the *logic* of fair betting quotients; these ideal reasoners can be assumed to be able to make all the relevant judgements. In order to make this clearer, return to the case of deductive logic.

¹³² However, see also the quote in Howson & Urbach 2004, p. 7, note 4.

¹³³ Note also that this problem persists even if intervals are used instead of point values, due to the exact cut-off points of the interval: see Howson & Urbach 1993, p. 88.

There, a set of propositions is said to be “consistent” if and only if it satisfies certain syntactic properties, i.e. it does not allow the deduction of a contradiction (see Enderton 2001, p. 119). Similarly here: a set of fair betting quotients is said to be “coherent”¹³⁴ if and only if they cannot be used in some combination to create unfair bets (i.e. to make Dutch books). In fact, the connection is even stronger: there are well known links between syntactic properties like consistency and semantic ones like satisfiability (in fact, these links are established by the celebrated and well known soundness and completeness theorems of standard mathematical logic: see Enderton 2001, section 2.5). In this way, one could map propositions truth functionally to the set of truth and falsity, and code the latter as 1 and 0. It then becomes immediately obvious that both standard propositional logic and the current treatment of fair odds can be seen to ultimately rely on the relationships between sets of numbers (see Howson & Urbach 2004, p. 15-16).

Return to the objection above. In mathematical logic, there is no expectation that some actual human agent can actually check the consistency of a set of propositions, or even be able to assign truth-values to the various propositions. This, however, does not mean that logic does not tell something about the relations between propositions an ideal agent could ascertain.¹³⁵ Similarly here: just because *actual* agents might not always be able to name fair odds, this does not mean that fair odds cannot be used to represent the beliefs of an *ideal* agent (see also Howson & Urbach 1993, p. 88-89).

¹³⁴ There is a terminological issue here as well: in order to mark the above ‘inductive’ logic from the standard deductive case, de Finetti and others have proposed to use ‘coherence’ in the place of ‘consistency’. Given the above connections between the two fields, this might however be somewhat unnecessary: see also de Finetti 1937, p. 93-94, Howson & Urbach 2004, p. 14 and Gillies 2000, p. 59.

¹³⁵ Equally, if ‘truth’ is seen as a metaphysical notion, then an actual agent’s inability to assign truth values to a proposition does not invalidate a Tarskian truth-definition: see also Enderton 2001, p. 83.

Moreover, it does not even mean that the above approach is not useful as a heuristic device for these actual agents: for example, it might be used to check the robustness of their reasoning under small changes. If some of my degrees of belief depend crucially on a small range of other degrees of beliefs, then it might be worthwhile for me to make sure that I feel confident that these other degrees of belief are indeed in this interval. These sorts of conclusions can be brought out of Howson & Urbach's system; in this sense, it is also at least somewhat a matter of pragmatic rationality to have degrees of belief that satisfy the axioms of the probability calculus: violating them might skew an agent's entire system of beliefs, thus making her less well equipped to handle the epistemological challenges the world holds for her.¹³⁶

One final word is in order about the brief arguments concerning statistical practice made above. Clearly, much more about this could be said about this here, also in defence of classical statistics.¹³⁷ However, since the present thesis is not concerned with questions concerning statistics as such, I shall not treat these issues here. What is to be kept in mind here is just that any probabilistic representation of beliefs has to have consequences of this sort for the methodology of the sciences. This will also be greatly important in the next three chapters.

5.3. Conclusion

The above account of Howson & Urbach's representation of beliefs stressed its concern with the *logic* of beliefs (rather than with issues concerning the pragmatic

¹³⁶ This justification of the pragmatic rationality of having beliefs that allow probabilistic representations adds a pragmatic element to the non-pragmatic Joyce 1998. See also above and section 6.2.

¹³⁷ A good source of arguments to this effect is Mayo 1996.

rationality of a probabilistic representation of beliefs) and with its seeking to connect this to a standard logic of *propositions*.

This linkage is also where both the main strengths and weakness of the account come from. On the side of the strengths, it might be seen to get the best of both worlds: it avoids the difficulties of preference-based accounts, and on the other, it avoids the shortcomings of treatments based on actual bets. On the side of the weaknesses, this combination once again walks a tightrope of logic and pragmatic issues that might lead to its falling into the crevice between what is desired and required of the representation of *actual* beliefs versus what is required and desired of *ideal* beliefs (for example concerning their notion of the ‘fairness’ of a bet).

What this means is merely that this representation – just like all of the other ones – comes at the theoretical cost of some conceptual worries and difficulties. However, having to pay a price does not necessarily mean that the incurred losses are not worth the benefits from the purchase.

Given the five accounts that span the literature in the subject presented here, it is now possible to take a step back and see where all of this leaves the probabilistic representation of beliefs. This is done in the next and final section of this chapter.

6. The Probabilistic Representation of Beliefs

At this point, it becomes possible to sum up and take stock of what this chapter has established when it comes to the cogency of the representation of beliefs by means of probabilities. This helps to give perspective to the various theories proposed and places them in a wider framework, but is also vital for the kind of work that will be done in the next two chapters: before one can launch an inquiry into the ways in which beliefs are *altered*, it is important to be clear about their *representation*. This is what this section aims at summarising; it starts by pointing out some communalities and differences amongst the theories, in order to gain a basic overview over this chapter. In the second section, this overview is analysed in terms of the distinction between logical and pragmatic rationality made earlier in this thesis. Finally, I come to an overall assessment as to what this chapter has shown concerning the feasibility of the probabilistic representation of beliefs.

6.1. The Structures of the Theories

The first thing to note here is that despite there seemingly being many different theories proposed concerning the representation of beliefs by means of probabilities, they can be grouped easily according to which of two broad directions they follow. On the one hand, there are preference-based accounts, on the other, betting-based ones.

The former start by assuming an agent has a preference ranking over a set of possible worlds (or acts, or propositions) that satisfies certain rationality constraints¹³⁸ and some

¹³⁸ As argued above, in the context of these theories, these “rationality postulates” should be understood as broadly relating to logical rationality.

technical assumptions, and then show that there is a function that is formally isomorphic to a probability function that captures the belief component of the preference ranking (they also show that there is another function that does the same for its desire component). The theories of Ramsey, Savage and Jeffrey are of this type.

The betting approach on the other hand begins by specifying a betting situation and then shows (possibly given a number of assumptions about the cultural, social and personal setting of the agent) that the agent's betting quotients can be used to measure her degrees of belief. Moreover, they show that these betting quotients satisfy the axioms of the probability calculus. The theories of de Finetti and Howson & Urbach fall into this category.

An important connection among the four *classical* treatments of this chapter (excluding Howson & Urbach 1993) is that they all rely on a causal theory of intentional action by assuming that an agent's beliefs and desires *cause* her to act in certain ways. De Finetti's betting approach uses this idea to set up specific situations that are structured in such a way as to allow easy recovery of the beliefs and desires from the actions (the bets). The preference approach relies on this theory in order to break the agent's preferences over possible worlds (or acts, or propositions) into a cognitive and desire-related component.

Seeing this connection between the two approaches makes clear that the preference approach can be seen to *encompass* the betting approach to some extent: the former deals with 'desires' in general, whereas the latter focuses on the specific desire for *money*. This is particularly obvious in the case of Ramsey and de Finetti: Ramsey's conditional prospects could roughly be construed as a more general version of de Finetti's bets (see

also Eells 1982, p. 65-66 and Gillies 2000, p. 54) – though it is important to always keep in mind that this might blur some important distinctions in Ramsey’s work.

The one treatment that stands out from this connection between the two approaches is that of Howson & Urbach, since they do not rely on the causal theory of intentional action. Their connection to the other works is in the form of relying on the previous formal results of de Finetti and combining them with a different philosophical basis (which however bears some affinities to Jeffrey’s treatment).

6.2. The Rationality of Probabilistic Representations of Belief

Given these connections between the theories, it is now possible to move towards an assessment of them. The first thing that has to be noted in this context is that none of the theories presented above is flawless. However, more important than noticing that all of them face some difficulties is to be clear about what these difficulties entail for the question of the possibility of probabilistic representations of belief. For example, Ramsey’s rather technical difficulties with his “ethically neutral propositions” were not seen to be greatly important for the overall cogency of his theory. On the other hand, Savage’s failure to be able to account for state-dependent utilities cuts much deeper into his account. What this section therefore aims at showing is the extent to which these theories (combined and individually) can be seen to provide arguments as to whether beliefs can be represented probabilistically, given that they all face some problems. The distinction between logical and pragmatic rationality made in chapter I turns out to be very useful in this context.

The first conclusion that can be drawn here is that for many agents, having beliefs that can be probabilistically represented is pragmatically rational. The fact that in some situations, Dutch books can be made against agents whose degrees of belief violate the axioms of the probability calculus is one argument to this score, at least for *some* agents in *some* situations. Another argument for the pragmatic rationality of probabilistic degrees of belief is the fact that agents whose beliefs are differently structured might have a very skewed picture of the world. This in turn might lead them to be less well adapted to their environment than they could otherwise have been, and thus increase the potential for making damaging decisions. For example, if an agent is unaware of the fact that some of her most cherished beliefs are dependent on yet further beliefs she is far less willing to endorse, then bringing out these connections can save her from making many disadvantageous decisions (not just in respect to losing money). This feature of the pragmatic rationality of probabilistic beliefs is particularly evident in the endorsement theories of Savage and Jeffrey, but it can also be derived from the treatments of Ramsey and Howson & Urbach.

In this way, having beliefs that obey the laws of probability will tend to be beneficial for the agent, as it helps her avoid these negative consequences. This in turn means that it appears to be very reasonable to assume that in some cases, and for some agents, degrees of belief should be represented by means of probabilities.

As made clear in the caveats above, this conclusion about the pragmatic rationality of probabilistically representable degrees of belief is not completely general: it holds in some cases only. This is due to the difficulties unearthed with the accounts presented in this chapter. However, complete generality is not to be expected from an argument about

pragmatic rationality. Clearly, there might be many situations where *no* harmful consequences accrue to agents whose beliefs violate the laws of probability.¹³⁹ What the argument from pragmatic rationality is meant to establish, however, is merely that there are *good grounds* for thinking actual agent's beliefs can be represented by means of probabilities. The fact that this conclusion is fallible only (and will have to be checked and re-checked frequently) is a direct consequence of this type of argument. In this way, the possibility of the probabilistic representation of beliefs cannot be *established* once and for all; however, these arguments provide a good *starting point* for the further investigation of this possibility.

However, some of the theories above have tried to go further than this by showing that there are also arguments concerning the *logical* rationality of having beliefs that are probabilistically representable. What the theories of Savage, Jeffrey and Ramsey have tried to establish here is that standard (deductive) logical arguments, together with the constraints imposed by these theories on the preferences of an ideal reasoner, are enough to make the latter one's beliefs probabilistically representable. On top of that, Howson & Urbach showed that it is equally the case that logical considerations are enough to ensure probabilistic representation of the ideal reasoner's beliefs if they can be assumed to make judgements about the fair odds of a bet.

What these arguments purport to show is that in principle, the very nature of beliefs entails that they are probabilistically representable: given sufficiently structured preferences or judgements about fair odds, the beliefs underlying those must obey the

¹³⁹ On the other hand, there also seem to be few conceivable negative consequences from having beliefs that *obey* the laws of probability (the ones that come to mind have to do with predictability in game theoretic circumstances; these sorts of issues are however too far removed from the discussion here, so I will not go into detail about them).

axioms of the probability calculus on pain of avoiding inconsistencies.¹⁴⁰ In this sense, there are grounds for thinking that having probabilistically representable beliefs is *logically* rational as well.

It has to be remembered at this point however that all the theories presented above have some problems in this respect; no theory was seen to be the panacea to all troubles. This fact might therefore be seen to call this last conclusion into question somewhat. That said, it should also be kept in mind that most of these theories can offer replies to their difficulties (with various degrees of cogency). This means that it is at the very least conceivable that they can be “patched up” eventually.

Moreover, even in those cases where the difficulties were seen to be very severe and the replies not fully successful in meeting them, some parts of the above conclusion might still be salvaged. The idea then would be to restrict the scope of the theory to the extent necessary to make it workable. This most likely will make that theory even more idealised and even less amendable to practical questions than it was before, but that might not be seen as a great loss, given the fact that the conclusions to be established are based on *logical* rationality anyway. For example, whilst it has been shown above that Jeffrey 1983 does not come to a unique representation of beliefs, it has also been argued that there are various ways of answering this objection. One of them notes that for cases in which one can place further constraints on the beliefs of the ideal agent, a unique representation of beliefs becomes feasible (see Joyce 1999, p. 138-145). On the other hand, it has been argued that Howson & Urbach’s approach faces the problem that fair odds might not always be plausible measures of an agent’s degrees of beliefs; one could

¹⁴⁰ If this avoidance of inconsistencies in one’s judgments or preferences can be seen to be an epistemological desideratum, then that will be enough to provide a further argument for the probabilistic representability of beliefs. This way of construing the argument is in the vein of Joyce 1998.

salvage this theory by claiming that it justifies a probabilistic representation of beliefs only in cases where fair odds are in fact good proxies of beliefs (e.g. for sincere agents without reservations about their opponents who can furthermore easily name what they take to be the fair odds for the situation in question). In this way, it seems plausible to assume that many of the problems can be taken care of at the cost of some added complications of the theory.

This means that on the whole, there are also good arguments based on logical rationality for the probabilistic representations of beliefs. These arguments have to be considered carefully due to the many problems with the various theories, but they can at least be seen to give some more support to the conclusion that beliefs can be represented by means of probabilities.

Before concluding with an overall assessment of what the final upshot of this discussion is, there is a last aspect about the distinction between logical and pragmatic rationality that is interesting to note. This aspect concerns the fact that the distinction between the two forms of rationality can both be used to *defend* a theory against objections as well as making it more *vulnerable* to them.

The distinction can be used as a *defence* against various objections, as being clear about what side of the divide a theory falls on allows it to disregard the other side to a large extent. For example, theories that are meant to show the logical rationality of probabilistically representable beliefs for ideal agents should not be called into question for failing to be straightforwardly applicable to real life. As was made clear in the context of Jeffrey 1983 and Howson & Urbach 1993, the limitations of actual agents are simply irrelevant to the aims of these theories.

However, at the same time, this distinction also makes a theory more *vulnerable* to objections: any theory that aims at a more comprehensive approach to the topic that includes the needs of both actual and ideal agents (and most theories fall into this category) stands in danger of conflating the distinction. Such a conflation though is to be avoided in order to maintain the overall cogency of the theory. For example, drawing conclusions about the logical rationality of probabilistic representations of beliefs from real life betting situations (as seen in the case of de Finetti) can be seem to be very unconvincing, as is the attempt to straightforwardly apply the conclusions drawn from assumptions about ideal agents to actual agents (as was seen in the case of Savage's worries with state-independence).

6.3. Summary and Outlook

All of this leads to the following overarching conclusion to be drawn: it seems on the whole to be safe to proceed by assuming that beliefs can be represented probabilistically. There are a number of convincing arguments as to the pragmatic rationality of having beliefs that are amendable to this treatment, though they are limited in terms of the situations they can be applied in. On top of that, there are a number of arguments at least pointing to the logical rationality of that issue as well, both concerning the nature of an ideal reasoner's preferences as well as her judgements concerning fair odds. In all therefore, the theoretical evidence amassed in this chapter points – with some reservations – to the general cogency of the probabilistic representation of beliefs.

One implication of this positive solution is that it is now possible to inquire into how beliefs should be and are *altered*. That is, it becomes possible to ask how beliefs are to be

revised, given that they are representable as probabilities. The particular revision at the forefront here concerns the case where new pieces of information are learned with certainty. As was also argued in chapter I, these kinds of considerations are not only important and interesting in their own right, but are furthermore at the heart of (the philosophy of) science. The rule for revising beliefs that is central in this context is Bayesian Conditionalisation, which thus forms the core of the discussion of chapters IV and V. The inquiry of this thesis will thus now move on to its third part, by discussing the properties and justifications of Bayesian Conditionalisation as a way of altering beliefs.

IV. Bayesian Conditionalisation and the Alteration of Beliefs: An Overview

The next task to be tackled in the argument of this thesis is to find a satisfactory account of how degrees of belief can and should be revised. Since the previous parts of the thesis have shown that there are good reasons to think that an agent's beliefs can be *represented* probabilistically, the focus of the rest of this inquiry is on probabilistic accounts of the *alteration* of beliefs.

In fact (as remarked earlier), the present inquiry largely limits itself to *one* account of the probabilistic alteration of beliefs¹⁴¹ only: "Bayesian Conditionalisation". This might seem overly restrictive: there are infinitely many ways of altering a probability function so that it still remains such a function – what makes Bayesian Conditionalisation so special?

Two arguments have been put forward in the literature to justify this pivotal position of Bayesian Conditionalisation: one based on the requirements of probabilistic representations of beliefs, and one based on 'dynamic' Dutch books. Understanding these arguments is imperative in order to grasp the scope and cogency of Bayesian Conditionalisation as a belief-revision rule. More than that though: the analysis given here is also the basis of the next chapter, which seeks to go beyond the justifications given in the literature so far to provide a novel argument for the *pragmatic rationality* of the probabilistic alteration of beliefs by means of Bayesian Conditionalisation.

¹⁴¹ Note that in what follows I use the term "conditionalisation" for this idea, i.e. the probabilistic revision of beliefs due to the receipt of new evidence.

A final remark is important here. Clearly, there are some fairly obvious connections between the alteration of beliefs and “scientific activity” (even though it might be far from easy to spell out exactly what these connections are). It is moreover the case that much of the debate about Bayesian Conditionalisation as a way of altering beliefs has been framed in terms of its applications in the philosophy of science.¹⁴² For ease of exposition and in order to make it somewhat easier to connect to this debate, I follow this general practice and base most of the illustrations (and indeed the general setup) of this chapter on issues in the sciences. This, however, should not be taken to mean that the questions of belief revision addressed here are limited to this domain – on the contrary, the arguments of this (and the next) chapter are completely general and attach to the alteration of beliefs *in general*.¹⁴³

The chapter is structured as follows: firstly, an overview of the most important properties of Bayesian Conditionalisation is provided in section 1. Secondly, the justifications for this rule are presented and analysed in section 2. Lastly, I conclude with an overall assessment of the justificatory state of Bayesian Conditionalisation in section 3.

1. The Soundness of Bayesian Conditionalisation

Bayesian Conditionalisation has two core ingredients: seeing beliefs as being representable by probabilities, and a particular way of altering these probabilistically

¹⁴² See for example Howson and Urbach 1993, p. 9.

¹⁴³ See also chapter I and chapter VI for more on the connection between the issues here and science.

represented beliefs. The former has been discussed at length in the previous chapters, so that the rest of this thesis will focus solely on the latter element.

To make clear how Bayesian Conditionalisation works, it is easiest to consider the case of a scientist who has to evaluate the support a new piece of evidence she has learned with certainty lends to the credibility of a theory or hypothesis. Her problem is to assess (given her initial degrees of belief in the theory and the piece of evidence) how exactly she is to alter her degrees of belief in this theory, now that the evidence has actually turned up. Intuitively, this is exactly what this conditionalisation rule allows her to do: it turns her previous (“prior”) degrees of belief into her new (“posterior”) ones.¹⁴⁴

The standard version of this conditionalisation rule can be stated as follows:¹⁴⁵

$$(BC) \quad P_{t+1}(H) = P_t(H/E) = \frac{P_t(H \& E)}{P_t(E)} = \frac{P_t(E/H) \cdot P_t(H)}{P_t(E)}$$

where H and E are the hypothesis and the piece of evidence in question, and the subscripts stand for various time periods: t is the period at the end of which the new evidence is known, and t+1 is the period afterwards. The presence of these time subscripts is also the crucial difference to the axioms of the probability calculus: the latter ones are synchronic only, whereas this rule links *multiple* time periods.¹⁴⁶ This means that the conditionalisation principle *as stated above* is an *addendum* to the (Bayesian

¹⁴⁴ An often conflated distinction is the one between “weight of support” - the “marginal” increase in the corroboration of a hypothesis – and the actual degree of corroboration; see also Gillies 1990. In general, it is crucial to keep them apart so as to avoid confusion; in the present context though, nothing hangs on the matter: this here is not an inquiry into confirmation theory.

¹⁴⁵ This statement is also implicit in Ramsey 1926, p. 87; see also chapter III. There are many other versions of this rule that are equivalent to the one above: see for example Howson & Urbach 1993, p. 28. However, for present purposes, the statement given here is the most useful.

¹⁴⁶ This is called the “dynamic assumption” by Hacking 1967.

reading of the) axioms of probability and does not *follow* from them (see also Howson & Urbach 1999, p. 99-100).¹⁴⁷

It can be easily shown that this rule is “sound” (i.e. that it satisfies the axioms of the probability calculus):

1. If H is a tautology (T) then

$$P_{t+1}(T) = P_t(T/E) = \frac{P_t(T \& E)}{P_t(E)} = \frac{P_t(T) + P_t(E) - P_t(E \vee T)}{P_t(E)} = \frac{P_t(E)}{P_t(E)} = 1$$

(as $P_t(T) = 1$ and $P_t(E \vee T) = P_t(T) = 1$)

2. If H is a contradiction (F) then

$$P_{t+1}(F) = P_t(F/E) = \frac{P_t(E/F) \cdot P_t(F)}{P_t(E)} = \frac{0}{P_t(E)} = 0$$

(as $P_t(F)=0$)

3. If H is a disjunction of two incompatible hypotheses H1 and H2 (so that $H=H1 \vee H2$ and $H1 \& H2=F$), then

$$\begin{aligned} P_{t+1}(H1 \vee H2) &= P_t(H1 \vee H2/E) = \frac{P_t((H1 \vee H2) \& E)}{P_t(E)} = \frac{P_t((H1 \& E) \vee (H2 \& E))}{P_t(E)} \\ &= \frac{P_t(H1 \& E) + P_t(H2 \& E)}{P_t(E)} = P_t(H1/E) + P_t(H2/E) = P_{t+1}(H1) + P_{t+1}(H2) \end{aligned}$$

(as $P_t((H1 \& E) \& (H2 \& E)) = 0$ as $P_t(H1 \& H2)=0$)

¹⁴⁷ This is an important point and will be picked up again in chapter V.

4. If H is conditional on some lemma (L) then

$$\begin{aligned}
 P_{t+1}(H/L) &= P_t((H/L)/E) = P_t(H/(L \& E)) = \frac{P_t(H \& L \& E)}{P_t(L \& E)} \\
 &= \frac{P_t((H \& L)/E) \cdot P_t(E)}{P_t(L \& E)} = P_{t+1}(H \& L) * \frac{P_t(E)}{P_t(L \& E)} \\
 &= P_{t+1}(H \& L) * \frac{P_t(E)}{P_t(L/E) \cdot P_t(E)} = \frac{P_{t+1}(H \& L)}{P_t(L/E)} = \frac{P_{t+1}(H \& L)}{P_{t+1}(L)}
 \end{aligned}$$

This soundness is important, as it shows that Bayesian Conditionalisation as a principle of belief revision does not change the nature of the representation of beliefs: if these beliefs could be represented by means of probabilities before the conditioning took place, then they remain representable in this way afterwards.

2. Bayesian Conditionalisation: Justifications and Objections

The above has shown that the conditionalisation principle is a sound rule and that *as stated above* it is an extra assumption and does not necessarily follow from a probabilistic representation of beliefs.¹⁴⁸ In terms of a *justification* for this rule, two kinds of projects should be distinguished: on the one hand, a justification for the *centrality* of rule BC could be provided (i.e. one that makes clear why *that* rule – out of the infinitely many possible ones – was chosen as the core account of the probabilistic alteration of beliefs). On the other hand, the aim could be a justification for the *cogency* of following rule BC, independently of its centrality (i.e. one that shows that Bayesian Conditionalisation is a

¹⁴⁸ It is important to note that this claim is true only for the way in which the principle was stated above; for more on this, see below.

“good” way of revising beliefs, which need not rule out the cogency of other conditionalisation rules at the same time).

It also has to be noted that these two different justificatory projects are not completely distinct from one another: showing the centrality of rule BC to all accounts of the probabilistic alteration of beliefs makes it at least plausible to also see it as a ‘good’ rule for revising beliefs. On the other hand, a very compelling justification for the rationality of Bayesian Conditionalisation *might* equally point to its centrality for any account of the alteration of beliefs.¹⁴⁹ Despite these connections, however, it is very helpful to keep these two kinds of justification apart (at least initially).

With this distinction in mind, I now present and analyse the two core justificatory accounts of rule BC in the literature. The first of these seeks to show the centrality and *logical* rationality of rule BC, whereas the second attempts to show the *pragmatic* rationality of using the rule.

2.1. The Cogency and Centrality of Rule BC: An Argument from Logical Rationality

The first justification to be discussed here tries to establish the cogency and centrality of Bayesian Conditionalisation by using an argument based on logical rationality. The structure of this justification is as follows: it firstly seeks to show that given a probabilistic representation of beliefs and the truth of two further assumptions, it is a matter of logical rationality that the revision of beliefs should follow the demands of rule BC. This first step thus shows that Bayesian Conditionalisation is a *cogent* rule (independently of its centrality). In the second step, the argument moves on to a

¹⁴⁹ This though will not be the case here.

justification for the *centrality* of rule BC by claiming that all of the necessary assumptions are in fact *always* true.

The two conditions that have to be added to the demand of a probabilistic *representation* of beliefs to get this argument for Bayesian Conditionalisation off the ground are the following (see Jeffrey 1983b, p. 80, Bradley 2005, p. 344 and Howson & Urbach 1993, p. 103-105):

1. (Certainty) The proposition to be conditioned on is learned with certainty:
 $P_{t+1}(E)=1$.

2. (Rigidity) Learning the proposition to be conditioned on does not require alteration of any conditional probabilities on it: $P_t (A_i/E) = P_{t+1} (A_i/E)$ for all i .

Both of these assumptions are extremely crucial for this justification and for the rest of this inquiry, so it is vital to be clear about what exactly they entail. It is further helpful to note that the following key difference between them: the first assumption is the main *distinguishing feature* of rule BC, whereas the second one is required in order for the rule to *work well* but is shared by other conditionalisation rules.¹⁵⁰ In other words, giving up the certainty assumption means giving up rule BC altogether (see for example Jeffrey 1983b, p. 80), but giving up rigidity means making use of the rule *problematic*, but no conceptual impossibility.¹⁵¹

¹⁵⁰ For example by Jeffrey Conditionalisation: see chapter V for more on this.

¹⁵¹ See chapter V for more on this.

In order to make it easier to distinguish the argument for the centrality of rule BC from that for its cogency later, it is useful to note that there are two readings of the certainty assumption, a strong one and a weak one.

The strong reading states that there is *always* such a proposition. That is, for every case of updating, a proposition can be found that was learned with certainty and which can be seen to be the reason for the updating (Glymour seemed to have been defender of this reading of the assumption: Glymour 1980, p. 69).

The weak reading sees the assumption as merely delineating the *scope* of the conditionalisation rule; it would then be seen as claiming that rule BC is applicable if and only if it is satisfied, without however there being a commitment to it *always* being satisfied (see for example Jeffrey 1983b, p. 81). Thus, there might be many cases where making the assumption would be false, which though merely means that rule BC cannot be used then. When the assumption is satisfied, though, Bayesian Conditionalisation *is* a good rule for the alteration of beliefs.

It is clearly the strong reading that is required for a defence of the centrality of rule BC, but since I discuss this point in more detail below, I shall not say more about it here. For now, I simply assume that the certainty assumption is satisfied, without distinguishing between the two ways of reading it.

The rigidity assumption on the other hand states that the proposition to be learned does not require alteration of any beliefs conditional on it. Much that can be said about it is similar to what has been said about the certainty assumption (e.g. whether there is always a proposition to be found that will satisfy it); in fact, as will be made clear below, there are many intimate connections between the two assumptions. An important difference for

present purposes though is that the rigidity assumption requires a vast amount of computational powers to be able to make sure it is satisfied: indeed, for sufficiently richly structured systems of beliefs, it would require computational powers that exceed those of any human agent (similarly to other aspects of deductive logic and probability theory: see Howson & Urbach 1993, p. 105 and Howson & Urbach 2004, p. 22-23). Checking whether the newly-acquired proposition requires the alteration of some (possibly very complex) conditional probabilities in the agent's web of beliefs is a very difficult – indeed, often impossible – task for any actual agent, though not of course for ideal reasoners. These latter ones *can* be assumed to be able to do this without any problems.

It is also important at this point to note that the two assumptions are interlinked: for example, failure of the certainty assumption can entail failure of the rigidity assumption.¹⁵² In this sense, any discussion of them really always needs to focus on both of them simultaneously. However, it turns out that all the points to be made in the present context can be made by giving more weight to the certainty assumption. An extra discussion of the rigidity assumption on top of this would merely add another layer of difficulty that is not fundamentally different from that discussed in the context of the certainty assumption.¹⁵³ Note that this is not to say that there are not also disanalogies between the two assumptions (e.g. only the rigidity assumption is diachronic), which might create *specific* problems for each of them. Rather, it points to the fact that all the problems that I shall raise here concern *both* of them, i.e. that the concentrating on the certainty assumption is enough to make all the points of interest here.

¹⁵² This can most easily be seen by considering an alteration of beliefs over the deductively closed set of two propositions A&B. Assume that $P_{t+1}(B)=s$, and $P_t(B)=r$, with $r \neq s$. Then (unless $P_{t+1}(A\&B)=P_t(A\&B)$) $P_{t+1}(A/B) \neq P_t(A/B)$.

¹⁵³ That said, I shall briefly drop the rigidity assumption in a case study in the next chapter. For a good overview of some of the effects of and reasons for rigidity failing to be satisfied, see Bradley 2005.

If the two conditions *are* satisfied, then it follows *from the axioms of the probability calculus* that updating of beliefs has to proceed according to rule BC.¹⁵⁴ This can be proven as follows.

Lemma: The only rule of belief change which satisfies the above assumptions and is such that it preserves the probabilistic representability of the new degrees of belief is Bayesian Conditionalisation (see for example Skyrms 1986, p. 190-191).

Proof:

Consider some updating procedure that yields a set of new degrees of belief in period $t+1$, and assume that both $P_{t+1}(A/E)$ (the result of the updating rule) and $P_t(A/E)$ are probabilities. Assume further that: $P_t(A/E) = P_{t+1}(A/E)$ (so that no degrees of belief conditional on E are altered).¹⁵⁵

$$\begin{aligned} \text{Then: } P_{t+1}(A) &= P_{t+1}(A/E) * P_{t+1}(E) + P_{t+1}(A/\neg E) * P_{t+1}(\neg E) \\ &= P_{t+1}(A/E) \text{ (as } P_{t+1}(E) = 1 \text{ by assumption, so that } P_{t+1}(\neg E) = 0) \end{aligned}$$

Thus $P_{t+1}(A) = P_{t+1}(A/E) = P_t(A/E)$ (due to the rigidity assumption).

$$\text{So finally: } P_{t+1}(A) = P_t(A/E) = \frac{P_t(A \& E)}{P_t(E)}$$

and the change must have happened by Bayesian Conditionalisation.

The upshot of this argument is the following: given the truth of the certainty and rigidity assumptions, if one's degrees of belief are meant to continue to satisfy the

¹⁵⁴ Technically, it also has to be the case that both $P_{t+1}(A)$ and $P_t(A)$ are greater than 0: see Jeffrey 1983b, p. 80. I will not discuss this extra condition at this point, however, as it does not add anything of importance here.

¹⁵⁵ I.e. assume rigidity holds.

mathematical laws of probability, they have to be updated using rule BC. This explains why rule BC is seen as so crucial: it can be proven that following it is quite simply a necessity of probabilistically represented beliefs (and the truth of the two assumptions above). In this way, obedience to it becomes a matter of logical rationality, since violating it would lead to an inconsistency with the requirements of the probabilistic representability of beliefs and the truth of the rigidity and certainty conditions.

It is important to note an implication of this derivation: if all that can be said in favour of rule BC is that it is a necessary consequence of the axioms and the rigidity and certainty assumptions, then it becomes *unnecessary* as an autonomous rule. In this way, the above argument could be seen to present another justification for the logical rationality of the *representation* of beliefs, but not for rule BC as an *independent* principle for their *alteration* (Howson & Urbach 1993, p. 103).¹⁵⁶ Thus, if this implication is taken seriously, it would mean that a discussion of rule BC on its own is simply misguided – this ‘rule’ turns out to be just another property of probabilistic representations of beliefs, in conjunction with the certainty and rigidity assumptions. This might be harmless when it comes to the *viability* of the rule, but it considerably lowers its *status*.

Leaving this issue aside though, it is key to note further that nothing has yet been said about the *centrality* of rule BC. In order for the justification here to show that Bayesian Conditionalisation is indeed a key ingredient in *all* accounts of the probabilistic alteration of beliefs, it has to be made clear that the certainty and rigidity assumptions are always satisfied. The other necessary assumption for the above derivation, namely that beliefs

¹⁵⁶ Note that when rule BC was first stated above, it was still an addendum to the probabilistic representation of beliefs as there was no reference made to the certainty and rigidity assumptions having to be satisfied.

can be represented probabilistically, was established to be relatively sound in the previous chapter, so that all that is left to be shown is the truth of these two assumptions. I shall further stipulate that the rigidity assumption is satisfied as well.¹⁵⁷ This thus means that (in terms of the distinction made earlier), reasons have to be given that establish the *strong* reading of the certainty assumption – that is, it has to be shown why it is reasonable to think the conditioning proposition is *always* learned with certainty. It is most useful at this point to consider an example.¹⁵⁸

Assume a zoologist is trying to establish whether a particular island is ecologically balanced, and that the presence of a rare kind of falcon is evidence to that effect. Assume further that one late November day at dusk, this zoologist observes at the far end of the island something that resembles this kind of falcon. Assume further that the rare species of falcon wanted as evidence for the hypothesis of ecological balance is difficult to distinguish in all but the best conditions (and dark November evenings are not amongst those) from a much more common species of falcon that is known to inhabit this island.

Given this scenario, a defence of the strong reading of the certainty assumption could be mounted in the following two ways. Either the example above represents a case where no proposition that is relevant for the conditioning is learned with certainty¹⁵⁹, or it does not. Start with the former case. If there is indeed no such proposition learned with certainty here, then (according to the strong reading) this means that this is not really a case of conditionalisation at all. Since the strong reading of the certainty assumption claimed that all genuine cases of updating involve the truth of the certainty assumption,

¹⁵⁷ As was made clear earlier, all the points that need to be made in the present context can be made with respect to the certainty assumption, so that I shall concentrate on this one here.

¹⁵⁸ What follows is adapted from Jeffrey 1983, p. 165-166.

¹⁵⁹ Really, this should read: “where no proposition is learned that describes all that the scientist has learned from the situation”. See below for more on this.

this must therefore mean (by a straightforward application of modus tollens) that this is not a genuine case of belief revision. If the scientist really cannot be sure about what she saw that evening, it might appear she has no grounds to change her beliefs at all.

The problem with this horn of the dilemma (for the above disjunction will turn out to be a dilemma) is that most people would intuitively disagree with the conclusion that *nothing* has changed with a view to the scientist's beliefs: the change might have been small, but it seems highly plausible that there was *some* change. For example, maybe she is now piqued enough to get up early the next day and wander to the other end of the island in the hope of seeing the bird again (something she would not have done without the prior sighting). This though entails that something *has* changed about her beliefs, which in turn makes it simply a very unintuitive (if not outright contradictory) conclusion to draw that no belief revision was *possible* here. In short: the above way of reasoning might be logically permissible, but it also appears to be highly counterintuitive.

This still leaves the second horn of the dilemma: a hard-line proponent of the centrality of rule BC might claim that there *is* a proposition that is learned with certainty, even in the above case (and in all cases like it): it might not be a simple proposition, but this does not mean that there is no such proposition. In fact, the argument of the previous reply can be turned around to say that, since it is uncritical that the scientist *did* alter her beliefs, and since rule BC is central to the alteration of beliefs, some proposition *must* have been learned with certainty. For example, it might be the proposition that the zoologist saw something that resembled the rare species of falcon.

The problem with this reply is twofold. Firstly, it is circular: it *assumes* the centrality of rule BC in an argument to this very conclusion. Secondly, it appears to be false

(overlooking the circularity). It might be possible to admit that there is *always* a proposition that is learned with certainty (even in cases of the above sort), but this does not entail that this proposition is also the one that is *conditioned* on: the scientist appears to have learned much more than the truth of *this* proposition.¹⁶⁰ Put crudely, the zoologist conditions on the possible sighting of a falcon, not on the certain sighting of a possible falcon (see also Jeffrey 1983, p. 166).

This means that on the whole, the argument for the centrality of rule BC has reached a dead end. Neither of the two possible ways of dealing with the above example appears to be very compelling. This will be even more obvious when contrasted with the other possible reaction to the example.

This other reaction claims (following Jeffrey) that there is indeed *no* proposition that is learned *with certainty* and describes *all the scientist has learned from the situation* here (see for example Jeffrey 1985, p. 67-68). It might indeed be the case that *some* proposition has been learned with certainty, but there is more than *this proposition* that can be gleaned from this situation. In essence, the arguments return here to Jeffrey's position of "radical probabilism" mentioned in chapter III: if certainty is simply no requirement for dealing successfully with the world around us, then the lack thereof also does not spell the end of the possibility of conditionalisation. To make this clearer, remember that Jeffrey held that probabilistic judgements are basic and not rooted in certainties at some fundamental level. If that is the case, then it should not be expected that there is always a proposition on hand to conditionalise on that is learned with certainty and represents the total of the experience undergone. This is also why the

¹⁶⁰ As will be made clear momentarily, this is the position often occupied by defenders of Jeffrey's radical probabilism.

second element of Jeffrey's radical probabilism is normally seen in this denial of Bayesian Conditionalisation as the only way to update beliefs (see for example Jeffrey 1983b, p. 78 and Bradley 2005, p. 344). The upshot of this though is that another conditionalisation principle is needed for those cases where the evidence to be conditioned on is not learned with certainty. This in turn implies that the centrality of rule BC has to be given up.¹⁶¹

Altogether, it therefore appears to be the case that there is little reason to assume that Bayesian Conditionalisation is central to all accounts of the alteration of beliefs. However, none of the above contradicts the *cogency* (in terms of logical rationality) of the rule as such: if the certainty assumption is read weakly as in the last reply to the biologist's story (so that it is seen to merely restrict the domain of applicability of the rule), then Bayesian Conditionalisation is still a rule that it is logically rational to follow whenever that assumption is satisfied. In this way, defenders of rule BC that are willing to give up the *centrality* of the rule can easily opt for the second reaction to the example above: they can claim that there are many instances where Bayesian Conditionalisation is not applicable (though other rules might be); crucially, though, they can claim that whenever it *is* applicable (i.e. where certainty and rigidity hold), it is logically rational to follow it.

What is crucial about even this weakened justification for Bayesian Conditionalisation though is that it provides nothing in the way of an argument for why *actual* agents should

¹⁶¹ Note also that in all of the above, the applicability of the rigidity assumption had been taken for granted. This however means that, in essence, the same argument could be re-run in terms of this assumption: it is possible to construct examples where changes in the relevant conditional probabilities are required, again ruling out the centrality of any principle that is based on it (like Jeffrey Conditionalisation): see Bradley 2005 and chapter V here. However, since denying rigidity would thus really only add another layer of difficulties to the present inquiry without presenting any new insights, I shall not pursue this point much further here (except for some brief remarks in a case study in chapter V).

use rule BC to update their degrees of belief: the justification above only concerns the *logical* rationality of doing so, not the *pragmatic* one. This is important, as this is one of the situations where the two come apart: the fact that it is logically rational to use rule BC if the two conditions are satisfied is irrelevant for the question as to whether it is also pragmatically rational to do so.

The reason for this is that (as was mentioned above), for any actual agent, there is no principled way of telling whether the rigidity assumption is satisfied or not. Only a Bayesian superhuman could check the deductive consequences of learning E together with one's background knowledge, which is a mammoth task that a real human being could never complete (see also Howson & Urbach 1993, p. 104-105). This though means that following rule BC cannot be justified as a matter of pragmatic rationality in this way. All the above justification for rule BC can establish (assuming the truth of the certainty and rigidity assumptions) is that it is logically rational to use it. If an argument for the *pragmatic* rationality of doing so is wanted, it has to be sought elsewhere.

2.2. The Dutch Book Argument Revisited: A Justification for the Cogency of Bayesian Conditionalisation based on Pragmatic Rationality

This pragmatic justification could be seen to lie in a return to Dutch book arguments: as noted in the context of the discussion of de Finetti 1937, these kinds of arguments are fundamentally based on considerations of pragmatic rationality. Moreover, it is very easy to set up a Dutch book argument for using rule BC.¹⁶² The general gist of this argument is that if you pre-announce and use a different conditionalisation principle from rule BC,

¹⁶² Pithy statements of this can be found in Howson & Urbach 1993, p. 100-101, Earman 1992, p. 46-47 and Bradley 2005, p. 357. The proof given here follows Skyrms 1986, p. 192.

then you can be made to bet at two different rates on the occurrence of the same event (or the truth of the same proposition). An intelligent opponent can use this to make a sure gain. To see this, consider the following argument.

Assume an agent announces some updating rule different from Bayesian Conditionalisation for the receipt of the news that proposition E is true. Assume further that this updating results in a new set of degrees of belief $P_{t+1}(A)$ for every proposition A. There are two possible cases (where, without loss of generality, A is some proposition the agent has beliefs over):

Case A: $P_t(A/E) < P_{t+1}(A/E)$

The bookie bets first on A conditional on E, and then against A conditional on E, with the agent setting conditional betting quotients¹⁶³ q_t and q_{t+1} for her conditional degrees of belief P_t and P_{t+1} (so that $q_t < q_{t+1}$).¹⁶⁴ The bookie chooses the stakes of the conditional bets as follows:

$$S_t > 0$$

$$S_t = S_{t+1}$$

Assuming E is true, there are two cases to consider:¹⁶⁵

a. A and E are both true.

The bookie gains

$$-[(1-q_t)S + (1-q_{t+1})S] = -(q_t - q_{t+1})S.$$

This is positive since $S > 0$ and $q_t < q_{t+1}$.

¹⁶³ For the definition of a conditional bet, see chapter III and the discussion of de Finetti 1937.

¹⁶⁴ Note that the bookie betting *on* A means that the agent bets *against* A. Furthermore, if q is the agent's betting quotient for a bet *on* A, then (1-q) is the agent's betting quotient for a bet *against* A. In case A occurs, the agent's "gain" is thus $-(1-q)S$.

¹⁶⁵ Remember that the bet is called off and all stakes are returned in case E is not true.

b. A is false and E is true

The bookie equally gains

$$-[(1-(1-q_t)S) - q_{t+1} S] = -[q_t S - q_{t+1} S] = -(q_t - q_{t+1}) S$$

Case B: $P_{t+1}(A/E) < P_t(A/E)$

By an argument exactly analogous to the one for case A, the bookie bets first against A conditional on E, and then on A conditional on E, this time setting stakes $S_t < 0$ and $S_t = S_{t+1}$. This will equally *guarantee* her winnings of $(q_t - q_{t+1}) S$

Define G to be the bookie's guaranteed winnings in case A or B. To obtain a Dutch book, the bookie finally simply bets some fraction S_f of G on $\neg E$ (where the agent's betting quotient on this simple bet is q_E):

$$S_f = \frac{G}{X}, \text{ with } X > 1.$$

In this way, she is guaranteed a winning amount, no matter what happens: in case E is true, she loses in the simple bet, but gains in the conditional bets. Thus, her winnings are

$$-[(1-q_E)\frac{G}{X}] + G = G(1 - \frac{1-q_E}{X}) = G(\frac{X - (1-q_E)}{X}).$$

As $X > 1 > (1 - q_E)$, this will always be positive. In case E is false she gains $q_E \frac{G}{X}$ (which is also positive), as the conditional bets on E are simply called off (Skyrms 1986, p. 192). This result can only be averted if $q_t = q_{t+1}$, in which case the updating proceeds according to rule BC. Thus, the argument concludes, the only way to avoid a ("dynamic") Dutch book is to use rule BC to update one's degrees of belief.

The main driving force behind this result is the fact that the agent is made to bet twice on the truth of the same proposition. This is being accomplished due to the fact that $P_{t+1}(A/E)$ and $P_t(A/E)$ are about the same proposition, so that any agent that announces betting quotients for them that differ is stand to lose. The analogue situation could be seen as that of first selling a good for $\$x$ and then buying it back for $\$(x+k)$ (where both x and k are greater than zero).

Crucial for this Dutch book result is the fact that the agent *pre-announces* her conditionalisation rule. This is the equivalent of her *agreeing explicitly* to betting at two different rates on the same proposition. It is very important to be aware of the importance of this fact for the derivation of this Dutch book result: if the agent did not *state beforehand* how she was going to update her degrees of belief, there would be no possibility of a Dutch book: the bookie simply would not know how to assign the stakes to the two bets. The bookie decides on the stakes she sets according to whether $P_t(A/E) > P_{t+1}(A/E)$ or $P_t(A/E) < P_{t+1}(A/E)$, which is only possible if the agent has announced these betting quotients beforehand.

This requirement of pre-announcing a conditionalisation rule is crucial here, as it is the main reason for why this Dutch book argument cannot be seen as successful in establishing the pragmatic rationality of Bayesian Conditionalisation. Before I come to this though, it is useful to mention two further difficulties for the argument above.

First, all the problems that attach to Dutch book arguments in general attach to this justification also, probably even more strongly. Clearly, if one was not convinced by the Dutch book approach to justifying a probabilistic representation of beliefs (as seems to be

true for most current Bayesians; see Howson & Urbach 2004, p. 9), it is unlikely that one would be convinced by the Dutch book approach here (see also Earman 1992, p. 49-50).

Moreover, even if one *did* defend the Dutch book argument for the representation of beliefs, this need not commit one to also defending it here. For instance, if one thinks it reasonable in certain restricted situations to use simple bets to elicit beliefs, this does in no way mean one also has to be convinced that conditionalisation is one of these situations.¹⁶⁶ The betting situation here is even more dubious than the one set out when finding a representation of beliefs, since it involves a rather complex *sequence* of bets, which might make the procedure even more prone to problems of risk aversion (and various other cultural, social and personal influences) than the one there.

Secondly, a version of this argument could be used to show that it is *never* rational to change one beliefs: changes in belief make it possible for a smart bookie to set up a betting scheme similar to the one above that allows her guaranteed winnings – except for the case where the agent does not alter her beliefs at all. This is clearly an untoward corollary of the theorem, and one that points to something that is deeply wrong with it.

However, even if these problem did not apply with full force or if a reply to them could be found, there is another, fatal problem for the above Dutch book argument for Bayesian Conditionalisation. To see this, it is best to briefly return to the general structure of arguments based on pragmatic rationality: these kinds of arguments are meant to show that the consequences of acting in a *particular* way (in real life) are advantageous (e.g. avoid guaranteed losses of money).

The problem with this for the above justification of Bayesian Conditionalisation is that this not at all what Dutch book argument shows. Instead, the pragmatically rational

¹⁶⁶ See also the discussion of Dutch book arguments in chapter III.

answer to the threat of a Dutch book for agents who do not use rule BC to revise their beliefs is not to *pre-commit* to an updating rule: as was seen above, this pre-commitment is the driving force behind the argument, without which it would collapse.

In this way, all that the argument can be seen to show is that it is pragmatically rational for an agent to refuse to play the bookie's game by simply keeping quiet about how she will alter her degrees of belief in the face of new certain evidence.¹⁶⁷ Crucially, this means that she is by no means forced to use rule BC on pain of pragmatic irrationality. The best way to avoid bad consequences is by refusing to bet in such a way as to allow a bookie to get a hold on the relevant betting quotients for purposes of a Dutch book – one is then free to update beliefs as one sees fit.

If this argument is found cogent, the Dutch book justification for rule BC must be seen to be very unconvincing: it relies on a false assumption about the pragmatic rationality of the agents in question. Moreover, even if one maintained that, despite this, the dynamic Dutch book argument still shows that it *is* pragmatically rational to use rule BC to revise beliefs, one will at least have to concede that this is not the only pragmatically rational strategy. The agent might also refuse to name her updating rule beforehand as an alternative way of avoiding the Dutch book. This makes using rule BC only one of *two* pragmatically rational ways of responding to the argument. Since it seems difficult to claim that the reply that advocates using rule BC is a priori preferable (if anything, there seems to be an a priori tilt towards the side of not pre-committing), this leaves the argument far short of being convincing.

¹⁶⁷ This is a well known point in the literature; for example, see Howson & Urbach 2004, p. 27, Earman 1992, p. 47 and Bradley 2005, p. 358.

The upshot of all this is that the dynamic Dutch book argument fails to provide a justification for the pragmatic rationality of Bayesian Conditionalisation. Since the first justification presented above has not done that either, no argument to that effect has been given up to now. Such a justification however might be very useful: giving actual agents – be they scientists or not – a helpful guiding principle with which to change their beliefs is very desirable. Bayesian Conditionalisation (at least on the face of it) seems to have something to offer here: despite the lack of convincing arguments in the literature so far, it does seem a very plausible and intuitive principle. If this plausibility could be spelled out in such a way as to provide a secure foundation to establish the pragmatic rationality of this rule, much might be gained. This task is the focus of the next chapter.

3. Conclusion

In this chapter, I have shown that Bayesian Conditionalisation is a sound rule of updating beliefs, and that there is a relatively compelling argument to show that it is a logically rational principle of the revision of beliefs *given the satisfaction of the certainty and rigidity assumptions*.

However, I have also argued that rule BC faces severe justificatory deficits when it comes to its *pragmatic* rationality. Neither of the two most common arguments to justify its use has been able to do so for actual agents with computational limitations.

What was therefore seen to be necessary is to provide a pragmatically rational justification for the cogency of Bayesian Conditionalisation not based on the rigidity

assumption or on betting-based considerations. Such a justification is presented in the next chapter.

V. Coherence in Thinking: A Pragmatic Justification for Bayesian Conditionalisation

The aim of this chapter is to provide a novel justification for Bayesian Conditionalisation: one that is based on an argument about *pragmatic* rationality. A justification of this sort is needed to make rule BC a compelling principle of belief-revision for *actual* agents, something that clearly is desirable. However, as shown in the previous chapter, in order to be compelling, a justification of this kind must not rely on the rigidity assumption (which was seen as too taxing on the computational powers of actual agents) or on betting (which was seen not to be compelling either). Instead, the novel justification has to be able to stand on its own, independent foundation.

However, since such a defence of the pragmatic rationality of Bayesian Conditionalisation implies seeing the rule in a very different light compared to a defence of its logical rationality (such as the one given in the previous chapter), the nature of the problems that will have to be tackled here is very different from that of those in the previous chapter. The best way to bring out these differences is through raising objections against the justification offered here. In particular, I present a detailed case-study of a potential worry to illustrate exactly what this defence of Bayesian Conditionalisation amounts to.

The structure of this chapter therefore is as follows. Section 1 presents the novel justification of Bayesian Conditionalisation as one pragmatically rational updating rule (out of possibly many more). Section 2 discusses some objections and replies, and presents a case-study of the potential difficulties of a failure of the rigidity assumption. Finally, I conclude in section 3.

1. Coherently Structured Thoughts: The Pragmatic Rationality of Bayesian Conditionalisation

As seen in the last chapter, the key to a justification for rule BC based on pragmatic rationality is in not relying on the rigidity assumption, or any other principle that is too taxing on the computational powers of actual human agents. Moreover, appeals to Dutch books (dynamic or otherwise) are also to be avoided. What is therefore needed is a novel justification of Bayesian Conditionalisation that is based on considerations of pragmatic rationality, yet steers clear of these pitfalls of the previous attempts.

In order to approach this new justification, it is best to begin by going back to the basic formulation of Bayesian Conditionalisation, and reconsidering exactly what it was meant to express. In its canonical formulation, the rule states that an agent's revised degrees of belief should be equal to her conditional degrees of belief before the revision took place, where the conditioning takes into account her initial degrees of belief in the truth of the proposition describing the evidence prompting the revision.

It is useful to illustrate this interpretation of rule BC with the following chain of reasoning: assume you are asked how likely you consider some hypothesis to be true, *assuming the truth of some proposition describing a certain piece of evidence*. Moreover, this evidential proposition is such that it does not require a shift in *reasoning* about the world. Right afterwards, you are told that this proposition actually does hold true, and

you fully believe its truth.¹⁶⁸ Would it not be entirely unreasonable *not* to set your new degree of belief in the hypothesis equal to the conditional degree of belief before?

Arguments concerning probabilistic reasoning in general are notorious for being treacherous when it comes to justifications by means of *intuitions*¹⁶⁹ – however, in this case, there surely does seem something imminently plausible about the above chain of reasoning. At least in certain straightforward cases, it seems completely compelling to set one’s new degrees of belief equal to the previous conditional ones. In fact, it is worth remembering that even Ramsey was convinced by this reasoning: as briefly mentioned in chapter III, he argued that it is “obviously” the case that rule BC should be followed (Ramsey 1926, p. 87). If it were possible to capture this intuition in the appropriate way, then that might be enough to place Bayesian Conditionalisation on a sound footing of pragmatic rationality.

This is exactly what I shall try to do in what is to follow: I seek to spell out in more detail where the plausibility of this way of looking at the rule comes from. As was seen in the previous chapter, in order to make an argument for rule BC that is completely general and unassailable – at least from a logical point of view – it has to be supported with the certainty and rigidity assumptions. Since this is not the route taken here, it should be noted from the outset that the justification presented here *cannot* be so completely general and exception-less. Making clear in what sense this is problematic (and in what not) thus forms a crucial part of the discussion of this chapter.

The main idea by way of which I mean to bring out the plausibility of following rule BC is that using this rule *structures* the agent’s thoughts in a particular manner, leading to

¹⁶⁸ The last clause is really the most important one: the issue under consideration is the updating of beliefs *in the light of further beliefs* only.

¹⁶⁹ A point that Gillies 2000 (p. 67) agrees with.

a specific, beneficial kind of coherence amongst her *decisions*.¹⁷⁰ It is this fact that will be used as the foundation of the novel justification for Bayesian Conditionalisation. However, before it is possible to argue for the *pragmatic rationality* of having coherently structured beliefs in this way, it has to be made clearer what exactly this means. That is, it has to be brought out clearly *how* rule BC manages to structure an actual agent's thinking about the world she lives in.

In more detail, there are two tasks to be accomplished here. Firstly, the justification for rule BC presented here has to show that is pragmatically rational to have *rules* for updating beliefs, instead of just altering beliefs in an ad-hoc manner. Secondly, it has to be made clear in what sense it is a sensible policy to use the *particular* rule 'Bayesian Conditionalisation' to do so. Dealing with these two tasks is the core aim of the next two subsections.

1.1. A Coherent System of Beliefs

The way rule BC achieves coherence in thought is by linking an agent's beliefs about the world at period $t+1$ to the beliefs in period t in such a way as to present an immediate connection between these two periods: her beliefs in the truth of some proposition A tomorrow are linked to her beliefs in the truth of A today, *where she assumes that her expectations of the world tomorrow turn out to be true*.

To express this slightly differently: if $P_t(H/E) = P_{t+1}(H)$, then that means that the agent's beliefs in the world of tomorrow (which differs from that of today only by containing one more piece of information: proposition E) depend on how that agent

¹⁷⁰ Here and in what follows I shall use 'coherence' in a non-technical, intuitive sense that is to be distinguished from that set out in chapter III.

thinks about H today if she assumes that E would actually be the case. That is, the agent reasons today *as if* proposition E were true, and tomorrow *since* it is true. In this way, the fundamental structure of the *hypothetical* world¹⁷¹ of tomorrow is identical to that of the *actual* world today – the only difference being that the former contains one extra proposition that is believed with certainty. This allows an agent to deal coherently with the world around her, by not prompting huge shifts in beliefs over time, but allowing for their smooth progression. In essence, an agent who revises her beliefs using Bayesian Conditionalisation ‘unifies’ her hypothetical reasoning and her actual reasoning: she identifies her beliefs about a hypothetical world (containing the hypothetical truth of an extra proposition) with her beliefs about an actual world (where that proposition *is* true).

Clearly, there are also many other ways of linking the agent’s beliefs about the hypothetical and the actual world; however, I want to argue that Bayesian Conditionalisation occupies something like a ‘default position’: if all an agent has learned is the truth of some proposition, and if she sees no reason to alter any of her beliefs that do not depend on E in any way¹⁷², then Bayesian Conditionalisation provides the primary way of linking these two ways of reasoning – at least until reasons appear to challenge these assumptions.¹⁷³ The reason for this status as a default position is in the intuitiveness of the principle I pointed to at the beginning of this section. Taking this intuitiveness as a fact of the alteration of beliefs means that rule BC gains precedent over any other rules that might be proposed. For Bayesian Conditionalisation to hold this privileged position

¹⁷¹ Note that I prefer to call this a ‘hypothetical’ as opposed to a ‘counterfactual’ world, due to the fact that at the time the revision of beliefs takes place, the hypothetical and the actual world might actually be identical (so that the hypothetical world is definitely not counterfactual). This goes back to the fact mentioned in chapter I that the present inquiry only deals with epistemological changes of belief.

¹⁷² This is simply because if $P_t(A/E) = P_t(A)$, then (by rule BC) $P_{t+1}(A) = P_t(A/E) = P_t(A)$. This also means that rigidity is satisfied: in this case, since $P_{t+1}(A) = P_{t+1}(A/E)$, it is also true that $P_{t+1}(A/E) = P_t(A/E)$.

¹⁷³ I thank Richard Bradley for this way of putting the argument.

does not mean that rule BC must always be *used*, but it does mean that it should at least always be *considered* as one of the possible alternatives.

It is important at this point to be very clear about the nature of this way of linking past and future beliefs: crucially, there is no *logical* contradiction in a wide divergence between hypothetical reasoning and actual reasoning. Consider some agent who thinks it is very likely that it is going to rain tomorrow if it is true that it is a very starry night tonight. There would be nothing *contradictory* in her then changing her mind to think rain very unlikely even if it was in fact a starry night (and she knows about this being the case).

What it *would* do, however, is jar the development of her beliefs over time: it would tear apart the beliefs she has to today from those she has tomorrow. Of course, the present justification is not to exclude the possibility of a change of mind on the part of the agent, or a fundamental change in what she thinks the world is like.¹⁷⁴ However, what it does want to exclude is the possibility of an agent *frequently* breaking with her previous thinking *even if* these circumstances are excluded (maybe because of a short time-span between the hypothetical and the actual reasoning, as argued below). Shifting her beliefs in this way in many situations would make it unclear in what sense she could be said to have a coherent character¹⁷⁵: her thinking at any one time would tend to be unrelated to her thinking at other times, thus breaking apart the cognitive links provided by a coherent character.

¹⁷⁴ This last might either refer to a metaphysical change in the world, or to a change in what the agent knows about this world, as argued in chapter I.

¹⁷⁵ Note that the above argument is meant merely as an illustration, and is not to be seen as a fully developed account of 'personhood': I am not trying to take a stance on the matter if violating rule BC in these circumstances means she does not have *any* character, or whether it merely means she has no *coherent* character. Since this issue is of no great relevance in the present context, the above illustration thus need not be specified further.

To see this more clearly, consider the following example: a scientist has just concluded an experiment. Before looking at a print-out of the data, she thinks that she will have found a great result (in terms of the paper she is going to write, say) if the data show a strong correlation between quantities X and Y, *ceteris paribus*. Immediately afterwards, she looks at the data and in fact finds this correlation. It would now appear very strange if she exclaimed ‘The experiment was a complete failure!’, *assuming she has found nothing else in the data* (as required by the *ceteris paribus* clause above).¹⁷⁶ Returning to the topic of the discussion: failure to use Bayesian Conditionalisation to alter one’s beliefs can lead to intuitively bizarre and outlandish ways of reasoning.

More than that though: having coherently structured beliefs in this way is also important from the viewpoint of the interpretation of others. It helps to ‘unify’ a person and allows others to make sense of her actions: it is harder, if not impossible, to gain an understanding of an agent’s motives and cognitive attitudes if these seem to subject to major shifts from one period to another and do not seem to bear resemblance to her motives and cognitive attitudes at a previous time period (excepting cases of a change of preferences or the way the world is¹⁷⁷). This means that assuming others use rule BC helps to make sense of their actions, just like using rule BC helps others to make sense of one’s own actions.

Again, many other rules might be seen to achieve this, but it is again the case that the intuitiveness of rule BC gives it a ‘default’ status: if it seems reasonable to assume that the other agent has learned nothing but the truth of some proposition and that she will not

¹⁷⁶ Of course, this is still not enough to completely rule out the possibility that she has changed her mind this very second or something else happened that was unforeseen. The extent to which this lack of guarantee is problematic for the justification will be discussed in more detail below.

¹⁷⁷ For some arguments as to the sense in which it is requirement of this justification that others are able to *tell* when a situation is an ‘exception’ or not, see below.

find it appropriate to change any beliefs not dependent on this proposition, then rule BC is to be seen at least as one of the contender of an interpretative principle for their actions¹⁷⁸: its intuitiveness and saliency make it plausible to assume that others will also be disposed to consider it one of the key updating rules available.

This can again be illustrated with the above case of the scientist. This time though, imagine that an assistant of hers is also present in the laboratory. Assume further that the scientist indeed exclaims her dismay at the results of the experiment, after having announced her happiness about the hypothetical result the second before. It now seems extremely likely that the assistant would be led to ask her *why* she reacts in this way? Is there something else the data have shown? Did she think of something just now that seriously puts the experiment in doubt, even in the presence of the correlation? Assume now that the scientist answers ‘no’ to all of these queries. In this case, it seems likely that the assistant would be at a *complete loss* to explain the scientist’s utterances. It is exactly this loss of being able to explain other’s thinking that rule BC helps to avoid.

In all therefore, I have tried to argue that Bayesian Conditionalisation should be seen as a compelling way of ‘uniting’ a person’s system of beliefs over time so as to ensure a more coherent interchange with the rest of the world (both in terms of the world as such as well as in terms of other agents), since it ties her hypothetical deliberations to her actual suchlike deliberations. In order to complete the justification of rule BC, it now needs to be established that it is *pragmatically rational* to have this kind of coherence in one’s set of beliefs.

¹⁷⁸ Once again, this will be the case until reasons appear that call this into question. For more on how damaging this lack of a guarantee of a successful application is, see below.

2.2. *The Pragmatic Rationality of a Coherent System of Beliefs*

The ultimate aim of the present justification for Bayesian Conditionalisation is to show it is pragmatically rational to use this rule for revising beliefs. In the first step, it has been shown that the crucial feature of the rule is its linking of the agent's cognitive attitudes over time in such a way as to help the agent to have a coherent character and to be able to interpret and be interpreted by other agents. The second step in the present justification therefore has to be the establishment of the fact that it is *pragmatically rational* to link one's cognitive attitudes over time in such a manner as to present a coherent whole.

Return to the definition of pragmatic rationality in chapter I: there, it was made clear that a principle is practically rational if it strongly tends to yield beneficial outcomes for the agent in question. In the present case, this means that it must be the case that having coherently structured beliefs in the above manner will equip an agent with cognitive tools that strongly tend to help her navigate the world in a better way than would otherwise have been the case. The following argument shows that this is indeed the case for rule BC.

To see this, return to the causal theory of intentional action of chapter III. The general gist of the argument to be presented here is that if an agent's desires and beliefs *cause* her actions, then having beliefs that are strongly variant are likely to cause her to make a number of disadvantageous decisions. However, if she uses Bayesian Conditionalisation to revise her beliefs, then her thinking tomorrow is connected to her thinking today in such a way as to avoid the kinds of jerks and jumps that cause problems in a largely stable environment such as ours. It means that altering beliefs and learning from the past

progresses in a controlled deliberate manner, and not in an over-excited rush or with dithering indecision.

In more detail, this can be spelled out as follows: consider an environment that is largely stable (i.e. which does not change much in short amounts of time), assume the time interval between t and $t+1$ to be short, and that the agent's desires remain largely constant over that interval. Such a situation represents the circumstances in which rule BC can be hoped to be applied without the presence of many disturbing causes. Given these conditions and the causal theory of intentional action, it follows that having beliefs that are structured by means of Bayesian Conditionalisation is pragmatically rational: it tends to keep the agent from making decisions that are damaging to herself.¹⁷⁹ In particular, it will help her avoid making decisions at two different time periods that are at odds with each other in such a way as to conspire to harm the agent.

The main path by which this harm can come about is the following: assume that at time t , the agent decides to do A if E were true. It is now important to note that taking a decision leads to the agent incurring costs – maybe due to her preparing to do A or due to her having to commit to doing A ahead of time. At the very least, she will have to incur the cost of the *deliberation* about what she would do if E were true – a cost that for example comes in the form of lost time that could have spent in other ways. If at time $t+1$ E actually turns out to be true and she fails to do A, the above costs all have to be counted as 'losses': since she ended up deciding against A, it would have been better not to have deliberated about what she would do beforehand, including her initiating the appropriate

¹⁷⁹ Note again that this is a qualified tendency claim: there are situations where, even given the above assumptions, using rule BC can be damaging to an agent. For example, agents with somewhat strange prior probabilities might update their beliefs in an unhelpful direction upon receipt of certain evidence. I thank Ken Gemes for helpful remarks on this point.

plans. In the best case, these losses from a useless deliberation at time t are all the losses the agent has to bear; in the worst case, the agent's decision at $t+1$ will involve yet more useless losses (if she changes her mind again at $t+2$).¹⁸⁰

A few words are in order here to make this deliberation-cost approach to the pragmatic rationality of rule BC clearer. Firstly, consider the above statement about the importance of a stable environment: this might initially appear to be a mis-construal of the question to be tackled. What is being inquired into are our *beliefs* about the world, not the world itself, so what does it matter *what the world is like*, when one is interested in our *beliefs* about it? This seems to be a confusion of metaphysics and epistemology.

However, it is vital to be aware of the structure of the above approach: it claims that using rule BC lowers the *costs* associated with decision making and facilitates the interpretation of others. These costs *are* connected to what the world is like – they provide the bridge between beliefs and the world. In this sense, this approach uses the same idea (though in a very different form) than de Finetti's Dutch book arguments: connecting concrete results to the status of an agent's system of beliefs.¹⁸¹

The second aspect of the above account worth pointing out in more detail is the distinction between "costs" and "losses": what the above account argues is that an agent whose belief alteration violates rule BC might have to face *losses* due to the *costs* of her decisions not being balanced by appropriate *gains*. That is, the agent incurs the costs of the deliberation without having the benefit of these deliberations leading to some helpful outcome.

¹⁸⁰ Note also that this is very different from the "sunk cost fallacy": in the latter case, an agent takes into consideration costs that cannot influence the object of her deliberation. Here, though it is the *presence* of these costs that is at issue, not whether they are taken into consideration.

¹⁸¹ The differences to the Dutch book approach will be made clear below.

This also links to the previous point about the stability of the environment: whether costs turn into losses in this way depends on a host of factors outside the agent's direct control. For example, a world that is subject to random shocks might penalise *following* rule BC and *reward* decisions that are at variance with one another across time: sticking to a plan might be more "costly" than reneging on one's previous decisions. What matters in these cases is the *balance* of the costs of keeping with one's prior intentions and those of changing one's mind. However, in those cases where the environment is stable (and the other conditions mentioned above are in place, like short time periods to decrease the likelihood of preference changes etc.) then Bayesian Conditionalisation is likely to lead to a reduction of the decision-making losses. This cannot be *guaranteed* to be the case – but it will tend to be so.¹⁸²

To illustrate this approach more vividly, return to the above example of the scientist in the laboratory. Assume this time that whilst the computer is calculating the correlations and preparing the print-out of the results, the scientist is asked by her assistant what work would have to be done if the experiment showed a high correlation between X and Y. By way of reply, the scientist might point to ordering in the results of other experiments, getting the paper written and formatted, and various tasks of this kind. Assume now that the assistant initiates some of this work whilst the computer is still calculating the results. If the scientist were then to reject her previous reasoning completely, even if the computer showed the desired high correlation, the assistant might justifiably be angry about the losses in time and money she had due to the scientist's indecision. Bayesian

¹⁸² For more on this point, see below in section 2.

Conditionalisation tends to prevent this situation from happening.¹⁸³ Note also that the scientist cannot always simply *refuse* to deliberate before the evidence is known: as in the example above, the outside world might *force* her to deliberate beforehand. Moreover, as argued below, not deliberating is also *unhelpful* for the agent, since it deprives her of the benefits of cooperation.

This example also shows clearly where the pragmatic irrationality of not using Bayesian Conditionalisation comes from: it lies in the fact that the beliefs will influence the agent to behave in a way that is self-defeating or harmful for herself. As made clear above, this is a generalisation of the philosophical lesson from the Dutch book arguments of the previous chapters: adherence to Bayesian Conditionalisation leads the agent to a coherent system of beliefs and thereby helps her avoid actions that ultimately come back to haunt her.

It is important though to also be aware of the *differences* between the approach here and the Dutch book argument: first and foremost, the above argument makes no appeal to “bets”, but attaches instead to the *general* consequences of having a coherently structured cognitive apparatus at one’s disposal. Secondly, it is not being claimed here that it is pragmatically rational to follow rule BC *because* it will avoid the possibility of Dutch books. Whilst it might clearly be the case that, in fact, using rule BC is connected to a potential avoidance of Dutch books, the latter is rather to be seen as the *consequence* of this pragmatic rationality of rule BC, and not as the *reason* for it.

Moreover, unlike the first justification of the previous chapter, the present account does not appeal to the rigidity assumption. On the positive side, this means that the latter

¹⁸³ Again, the tendency qualification is important: the existence of strange priors may interfere with rule BC actually having beneficial consequences.

becomes accessible to actual agents (as was made clear in the previous chapter); on the negative side, it means that it is far less clear when it is applicable. In case of the previous justifications, it was possible to specify precisely the conditions under which it was applicable. In the present case, even the appeal to the absence of prominent disturbing causes (due to a short timeframe of the revision, a stable environment etc.) is not enough to conclusively specify when use of rule BC is pragmatically rational. This, however, is entirely in line with what is to be expected from such a justification, but since I discuss this point in more detail below, I shall not say more about it here.

There is also a second argument for the pragmatic rationality using Bayesian Conditionalisation (paralleling the reasoning in the previous subsection) that it is important to mention here. This second argument is based on the fact that having coherently structured beliefs and assuming that others do so, too, is beneficial for one's dealings with other human agents. To see this, remember that rule BC works in two ways here: on the one hand, by being a helpful tool of one's interpretation of others, and on the other, by helping others to make sense of one's own behaviour. Both of these means of interpretation are highly beneficial in a world that is as dependent on large amounts of inter-personal cooperation as ours. This means that following rule BC is pragmatically rational as it opens up the possibility of communication, which in turn enables cooperation – and cooperation is beneficial to an agent.¹⁸⁴

To see this in more detail, note that Bayesian Conditionalisation allows agents to communicate and prepare for a future of cooperation more easily than might otherwise be

¹⁸⁴ There is an issue here as to the interplay between the possibility of communication and the possibility of cooperation; however, for present purposes, it is not necessary to delve deeper into this problem as it is enough here to note that there is a connection. Also, I shall side step completely all issues connected to game theoretic problems of cooperation and defection as they are not greatly pertinent in the present case.

possible. It allows a number of agents to come together and agree on what they would do if a certain circumstance were to obtain. If that circumstance then actually does obtain, they can put their plans into practice. Agents violating rule BC might have difficulties in taking part in a scheme of this sort (barring some other publicly accepted rule of belief revision like Jeffrey Conditionalisation), as they could not be relied on to follow through on their hypothetical reasoning (moreover, they might not want to take part in it, since they would not think that others would follow through on *their* hypothetical reasoning). Not being able to take part in such a scheme then makes the possibility of inter-subjective cooperation (which I clearly take to be beneficial here) much harder for these agents; positively speaking therefore, using Bayesian Conditionalisation is pragmatically rational as it enables various agents to come together and cooperate for their mutual benefit. This is also why not using rules at all to update beliefs is an inadvisable strategy to deal with the above case of losses from decision-making: without a rule-based updating procedure, one is incapable of taking part in and benefiting from intersubjective cooperation.

This can again be illustrated by means of the example of the scientist and her assistant. As was made clear earlier, one (intuitively appealing) way for the two of them to work together on the publication of the paper about the experiment is by taking each other to be using Bayesian Conditionalisation in the situation they are facing (at least in ordinary circumstances). Otherwise, they might find it more difficult to agree on the steps to take when the results are available; in effect, they might struggle to cooperate or plan ahead at all (in the absence of *any* obvious candidate rule for the revision of beliefs). This is highly detrimental to the success and range of their deliberations.

In this way, I have tried to show that revising beliefs by means of Bayesian Conditionalisation is pragmatically rational: it structures one's thinking about the world in such a way as to make coherent intentional actions possible. This in turn is beneficial in a largely stable environment with much social interaction, as is true of ours.

Before moving on to a closer critical scrutiny of this justification, it is important to note the following aspect of it. The justification for Bayesian Conditionalisation here given still depends on the satisfaction of the certainty assumption, in that the proposition to be conditioned on is learned with certainty. As noted in the previous chapter, however, this is due to the very nature of this conditionalisation rule. As soon as this assumption is not considered to be adequate anymore, a different conditionalisation rule has to be sought for. However, as in chapter IV, this is unproblematic for an argument as to the *cogency* of Bayesian Conditionalisation (as opposed to its *centrality*): since the argument here is merely meant to show the pragmatic rationality of *rule BC*, it is inconsequential whether there are other rules that are pragmatically rational in other circumstances. For example, in the case of the receipt of uncertain evidence (i.e. when the uncertainty assumption is not satisfied), it might be pragmatically rational to use Jeffrey Conditionalisation in order to change one's beliefs.¹⁸⁵ The justification given here claims merely that rule BC is *one* (out of possibly many) useful ways of coherently thinking about the world – it concedes that there are circumstances where other ways of updating beliefs might be more helpful than rule BC.¹⁸⁶ More about this will also be said below.

To summarise the above considerations, it is helpful to return to the two tasks set for the account presented here: showing why rules for the alteration of beliefs are

¹⁸⁵ This is even more compelling if it is noted that Jeffrey Conditionalisation collapses into Bayesian Conditionalisation for certain evidence. See below for more on this.

¹⁸⁶ In this way, it is obviously much closer to Jeffrey's take on conditionalisation than to that of Glymour.

pragmatically rational, and why using Bayesian Conditionalisation as a particular specimen of such a rule is so, too. What the above considerations have shown here is that, firstly, it is pragmatically rational to have rules for the alteration of beliefs since they structure an agent's beliefs in such a way as to cut down on the losses of making decisions and allow for cooperation with other agents. Secondly, the *particular* usefulness of rule BC in this comes from the fact that its intuitiveness gives a default position in the set of possible conditionalisation rules, and that in a number of circumstances, using it will tend to have beneficial consequences (for the same reasons that attach to conditionalisation rules in general). Together, these two arguments make up the new justification for the pragmatic rationality of using rule BC to be presented here.

2. Objections and Replies

It is at this point very helpful to consider some possible objections to this justification of rule BC, in order to see fully what it entails and what not. I shall also present a case study of some difficulties for Bayesian Conditionalisation as it is construed here, so as to clarify the nature of the justification defended here.

2.1. Two Objections and Two Replies

Firstly, one might claim that the above justification of rule BC is too conservative: sometimes it might be necessary to initiate large shifts in one's beliefs, and not follow the gradual process inherent in Bayesian Conditionalisation. Especially in circumstances where being too predictable or too limited in one's reactions to the world is damaging,

following rule BC might not be helpful at all. In the context of the philosophy of science, this can be expressed as follows (see also Earman 1992, p. 195-196 and Gillies 2000, p. 74): sometimes, a shift in the whole paradigm is required to push science forward. Small changes in belief cannot do this; a complete overhaul over one's system of beliefs might be necessary. For example, rule BC might be an appropriate way of updating beliefs in some hypothesis if the scientist is right about the general framework the hypothesis is set in. If, however, she is wrong about that, a much more radical shift in perspective becomes necessary (see also Gillies 2000, p. 74).

In more detail, this is to be understood as follows: there might be circumstances where the probability function representing one's beliefs has to be changed discontinuously and without regards to the limitations of the evidence that has been learned in a certain situation. Intuitively, there might be situations where the data suggest a minor alteration in a scientist's beliefs, but external considerations call for a complete *overhaul* of them.¹⁸⁷ Such a situation will occur whenever the rigidity condition is not satisfied: if it is the case that some conditional probabilities have to be altered due to the agent learning the truth of a proposition, no amount of Bayesian conditionalising will do justice to the situation.¹⁸⁸

In more general terms, this objection returns to the worry mentioned earlier about the limited scope of the above justification: if the environment the agent acts in is prone to quick changes, then linking one's decisions in the past to those in the future can be very unhelpful, as the world around the agent might have changed considerably. The same

¹⁸⁷ In a certain sense, this could be construed as taking this issue out of the domain of epistemology and philosophy of science and into decision theory and practical rationality: instead of asking "What theory fits the world better?", it might be better to ask "Which theory is it better to adopt?". I thank Richard Bradley for helpful comments on this point.

¹⁸⁸ This will be made clearer below in the case study.

goes for long time periods between the first deliberation (when the evidential propositional was only hypothetically taken into account) and the second one (when it was actually taken into account): the longer the period, the more plausible is it that the agent would have changed her mind or that the world has changed around her. This in turn would mean that having the kind of coherence offered by Bayesian Conditionalisation between past and future decisions might not be very helpful at all. In these cases, more radical changes of belief are necessary.

However, despite there being nothing wrong in the reasoning of this objection, it misses the mark nonetheless. The reason for this is that it misconstrues the nature of the pragmatic rationality of Bayesian Conditionalisation: the above objection shows that there are *exceptions* to the applicability of rule BC – this, however, is entirely unproblematic for a pragmatically rational principle. The justification above argues that Bayesian Conditionalisation *strongly tends to* have beneficial consequences – which is not to say that it will always *actually* have them. Pragmatically rational rules are limited in terms of the kinds of situations they are applicable in: they are not always applicable, they are not applicable in all places, and they are not applicable for all agents. This however is just what distinguishes them from logically rational rules, which are *in principle* always applicable (as was made clear in chapter I).

What the above objection shows therefore is that there are “disturbing causes” (in the terminology of Mill 1844) that might alter and cancel the strong tendency of rule BC to have beneficial consequences: for example, the presence of an unstable environment, or fundamental changes in the agent’s preference structure. Once again though, this merely means that rule BC is not completely reliable to actually yield beneficial consequences

(though it is completely reliable to strongly *tend* to do so). The upshot of this is that application of the rule has to be checked and other factors taken into consideration. Clearly, this does not distract from either its pragmatic rationality or its usefulness in practice: it just means that there is no guarantee of success – which though should not be expected from a pragmatically rational rule.

In order to be clear about what is being claimed here, bringing up a brief further objection might be helpful. It might be argued that this talk about tendencies and disturbing causes can only be metaphorical: rule BC is about the alteration of *beliefs*, so at most, using it might be said to yield the *right answer* to the demands of some situations – which is very different in nature than talk of ‘disturbing causes’. This latter way of presenting the situation thus might be seen to be highly misleading, as it gives a false sense of the structure of the problem: instead of talking about beliefs and their alteration, the debate has been shifted to talk about ‘causal mechanisms’: a category mistake of sorts.

The reason why this objection fails to be compelling is that it misconstrues what the talk about tendencies applies to. What was being claimed above is that in order to *characterise* the “right answers” in “some” situations, talk of tendencies is helpful. That is, when it is claimed that Bayesian Conditionalisation tends to have beneficial results, this should be understood as claiming that the kind of answers given by the rule (i.e. the kinds of alterations of belief initiated) tend to have consequences that are beneficial. It is also in this way that disturbing causes enter into the picture: it is not that they disturb the

workings of rule BC directly – instead they disturb the rule from having beneficial consequences.¹⁸⁹

This immediately leads to the second objection to this justification for Bayesian Conditionalisation. As was made clear in the previous chapter, checking if the certainty and rigidity assumptions are satisfied is practically unworkable. In a similar manner, it could be noted that there seems to be no principled way of *testing* whether the current situation is one where rule BC is to be applied, or if it is one of the “exceptional” circumstances in which countervailing factors cancel the beneficial tendency of the rule. This lack of certainty concerning its successful application in turn means that Bayesian Conditionalisation might be seen to be hardly useful at all: the fact that there are exceptions might be tolerable, but the fact that it is impossible to tell *when* something is one is not.

However, once again, it is important to keep in mind what it means to be a pragmatically rational principle. Pragmatically rational principles specify merely what they would bring about on their own, in the absence of any disturbing causes. It is not part of the pragmatic rationality of the principle to state the presence or absence of these disturbing causes: the rule merely states the *tendency*, not the actual outcome. Thus, whilst it is true that there are no hard-and-fast ways in any actual situation of making sure that there are no disturbing causes present, this is not something that in any way impinges on the pragmatic rationality of the principle.

Given this reply, the objection could of course continue by noting that in this case, it is problematic whether the pragmatically rational principle is at all useful in practice: it

¹⁸⁹ Making this distinction clear was also the point of splitting the discussion in the first section of this chapter.

might be true that the lack of certainty of when rule BC can be applied does not interfere with its *pragmatic rationality* directly, but it at least should be seen to make the *applicability* of the rule doubtful.

There are two points to be noted in reply to this. Firstly (as specified in chapter I), pragmatically rational rules show a *strong* tendency to have beneficial results: this is to mean that even the presence of many relatively weak disturbing causes will not hinder the tendency from achieving its beneficial outcomes. In the case of rule BC this might mean that even modest changes in tastes or the world need not completely hinder the workings of Bayesian Conditionalisation to beneficially structure beliefs.

Secondly, despite this, it *is* true that there is no clear-cut way of determining how many weak disturbing causes it takes to disrupt the workings of rule BC, or how strong a single disturbing cause has to be to do so. In practice, this means that rule BC (just like all pragmatically rational rules) truly is a fallible, heuristic device. It requires the user to be constantly on the alert as to possible signs of its misapplication and as to the possibility of mistakes.

This though is not as problematic as the objection makes it out to be: there are many situations where all we have at our disposal are pragmatic and fallible tools (one such candidate situation – science – is scrutinised in more detail in the next chapter). Being fallible however does not mean that these tools are not useable – it merely means that they have to be used with care. There are no guarantees one will always make the right decision, but this does not mean that one never will.

2.2. *The Failure of Rigidity and Certainty: A Case Study*

At this point, it is very useful to present a more detailed look into what can go wrong with the above justification for rule BC, and what that entails for the justification in terms of pragmatic rationality presented here. A particularly interesting case for this presents itself with a view to the previous chapter and the problems that were seen to come from too much reliance on the certainty and rigidity assumptions. Here, I shall look at this situation from the other side to present a “case study” of what can happen when rigidity and certainty are *not* satisfied. The failure of certainty is interesting on top of this also because it entails giving up rule BC for a different conditionalisation principle like Jeffrey Conditionalisation. This is important here firstly since it shows that Bayesian Conditionalisation is not *the only* pragmatically rational rule for the alteration of beliefs, but merely one amongst many (e.g. Jeffrey Conditionalisation, what Bradley 2005 calls “Adams Conditionalisation”, and a host of unnamed other ones). Secondly, it also makes clear how the kinds of issues addressed here have general importance, since they will also have to be tackled by justifications in terms of pragmatic rationality of other conditionalisation rules.

It is best to begin by considering what can happen if the rigidity condition is not satisfied.¹⁹⁰ To see this, consider the following example, loosely based on the idea of the ‘Kuhnian meta-induction’.¹⁹¹ Assume there is some hypothesis in whose truth a scientist

¹⁹⁰ Note that there is something awkward about an assumption failing to be satisfied that is not actually made. However, the way this is to be understood here is as follows: whether or not the rigidity assumption is actually explicitly introduced into the justification or not, it can be seen to *describe* the situation as either calling for the revision of the relevant conditional probabilities or not.

¹⁹¹ The following is an extended version of an example originally due to Howson & Urbach: see Howson & Urbach 2004, p. 27-28 and Howson & Urbach 1993, p. 101.

currently has full faith (say, the theory of evolution); let that theory be expressed in proposition H. Then it is the case that

$$P_t(H) = 1$$

Also, assume further that reading “The Structure of Scientific Revolutions” can convince a scientist that, eventually, even H will prove to be false in at least some respects (maybe the scientists are deeply impressed by the claim that this has been the case throughout history for every theory that has been proposed). That is, consider the proposition that “ $P_{t+1}(H) = s$ ” (which is meant to express this Kuhnian meta-induction); call this proposition K. Assume further that the scientist in question assigns some positive probability to the proposition that K will turn out to be true in her case, too, i.e.

$$P_t(K) = r$$

where $r \in (0, 1)$.¹⁹²

Assume now that the time has passed and that she has indeed finished reading Kuhn’s book – what should her new degrees of belief in the theory be? Clearly, by the axioms of probability:

$$P_t(H/K) = \frac{P_t(H \& K)}{P_t(K)} = \frac{P_t(H) + P_t(K) - P_t(H \vee K)}{P_t(K)} = \frac{1 + r - P_t(H \vee K)}{r}$$

¹⁹² Note that this does not entail (vs. van Fraassen 1984, p. 244) that the agent has incoherent beliefs, as will be argued below.

Now, note that as $P_t(H \vee K) \in [0, 1]$ it must be the case that $[1+r \cdot P_t(H \vee K)] \in [r, (1+r)]$.

Since $P_t(H/K) \in [0, 1]$, it must therefore be true that $P_t(H \vee K) = 1$. Thus:

$$P_t(H/K) = \frac{1+r-1}{r} = 1$$

This though means that her degrees of belief in H next period are (using rule BC):

$$P_{t+1}(H) = P_t(H/K) = 1$$

This is the consequence of the “once certain, always certain” property of the axioms: if one is certain about the occurrence of an event (or the truth of a proposition) then no amount of conditioning can lead one to waver in this judgement.

However, assume now that Kuhn’s arguments have actually convinced the scientist that H is not certainly true anymore, just as she thought possible (though not certain earlier on). It should then be the case that in fact

$$P_{t+1}(H) = s$$

which is a contradiction as $s \neq 1$ (by assumption also).

What this argument shows is the pitfalls of the failure of the rigidity assumption. The example above is predicated on the fact that some *conditional* probabilities will have to

be changed from time t to time $t+1$: it is the consequence of a probabilistic representation of beliefs that

$$P_t(H/K) = 1,$$

however, the structure of the argument requires

$$P_{t+1}(H/K) = s.$$

Since the whole point of Kuhn's book (in the present example) is to make the scientist *less* certain of the truth of H it seems ludicrous to argue that the conditional probability $P_t(H/K) = 1$ is an accurate representation of the scientist's beliefs. This though is a violation of the rigidity assumption, as it requires the alteration of a probability conditional on the proposition prompting the conditioning. This thus explains why the above example leads to a contradiction: the violation of rigidity is a disturbing cause of the beneficial tendency of rule BC and leads to the wrong structuring of the scientist's beliefs.

At this point, it is also important to note that this example cannot be discounted on the grounds that proposition K is a sort of "meta-proposition" – a proposition about a future degree of belief in another proposition – and that propositions of this kind should not be treated in the same way as 'ordinary' propositions. This is because K seems to in fact be a proposition like any other, and there is no reason why it should not be possible to have a

degree of belief in it, just like in any other proposition. Claiming anything else seems very ad hoc and unconvincing.

The example also cannot be discounted on the grounds that the scientist's beliefs before reading Kuhn's book are incoherent (they are clearly not), or that the grounds for the alteration of beliefs are not epistemically justifiable: it is not that reading Kuhn is like getting drunk or taking drugs: it is just that reading it requires alteration of some conditional probabilities.

Some more insights into the nature of the above justification can be gotten by also studying the effects of dropping the certainty assumption. This implies that Bayesian Conditionalisation must be given up: it is based on the fact that the proposition to be conditioned on is learned with certainty.¹⁹³ If that assumption is dropped, rule BC has to be dropped as well.

However, other conditionalisation rules have been introduced exactly for this situation; the most famous among them is Jeffrey Conditionalisation (see Jeffrey 1983, chapter 11). As was mentioned earlier, this is an important rule in its own right and thus deserves at least a brief discussion, even in an inquiry primarily focused on Bayesian Conditionalisation. Its introduction here is furthermore useful for showing that the justification given in this chapter does not single out rule BC as the only valid rule for altering beliefs, but that other rules are likely to face similar issues.

In general, Jeffrey Conditionalisation works as follows: assume E_i is the i th element of a set of n mutually exclusive and exhaustive propositions. Given this partition, an agent's new degrees of belief in some proposition H are derived from the old according to:

¹⁹³ The agent's degree of belief in the truth of that proposition can be less than 1 before the truth is revealed of course – the important point is that the *truth* will be revealed with certainty afterwards.

$$P_{t+1}(H) = \sum_{i=1}^n P_t(H/E_i) * P_{t+1}(E_i)$$

Note firstly that this rule collapses into rule BC for certain evidence (if one of the E_i is 1, the others must be 0). Secondly (and of great interest in the present case), Jeffrey Conditionalisation still relies on the rigidity assumption in order to be applicable (see Jeffrey 1983b, p. 80). This implies that the discussion of this and last chapter could without much change be transferred to Jeffrey Conditionalisation: there is a justification of its logical rationality based on the rigidity assumption (see Jeffrey 1983b, p. 80), there is a Dutch book argument for it (see for example Bradley 2005, p. 356), and it is conceivable that one could formulate a defence of its pragmatic rationality by basing it on its property of structuring thoughts in the face of the receipt of uncertain evidence.¹⁹⁴

A further implication of this is that one can continue the case study from above by showing what happens in case *both* the certainty and the rigidity assumptions fail to be satisfied. To see this, keep all of the above setup, except for setting

$$P_{t+1}(K) = t$$

In this case, Jeffrey Conditionalisation implies that:

$$P_{t+1}(H) = P_t(H/K) * P_{t+1}(K) + P_t(H/\neg K) * P_{t+1}(\neg K).$$

¹⁹⁴ This also shows why the restriction in this thesis to only consider rule BC as a way of probabilistically altering beliefs is not as limiting as it might at first appear: very similar problems to those unearthed here also affect many other conditionalisation rules (see also Howson & Urbach 2004, p. 29 and Howson & Urbach 1993, p. 105-110).

By the same argument about the ‘once certain, always certain’ property of the axioms noted above, it is still the case that $P_t(H/K) = 1$. Further, it therefore also holds that:

$$P_t(H/\neg K) = \frac{P_t(H \& \neg K)}{P_t(\neg K)} = \frac{P_t(H) + P_t(\neg K) - P_t(H \vee \neg K)}{P_t(\neg K)} = \frac{1 + (1-r) - 1}{1-r} = 1$$

This implies that

$$P_{t+1}(H) = (1 * t) + [1 * (1-t)] = 1$$

Once again therefore, reading Kuhn’s uncertainty-inducing book has no influence on the agent’s degrees of belief, just as in the case of rule BC above. Moreover, just as before, this is due to the violation of the rigidity assumption (or to be more precise: its non-violation where it should have been violated): the axioms imply that $P_t(H/K) = 1$, whereas the example requires that that $P_{t+1}(H/K) = r$.

Note also that one cannot discount this result on the grounds that it relies on some propositions being believed with certainty, and that this certainty is inadequate for a rule like Jeffrey Conditionalisation (as argued above, this conditionalisation rule is fundamentally rooted in Jeffrey’s radical probabilism¹⁹⁵). The reason for the fact that this argument fails is that the example above can also be run for the case in which *no* propositions are believed with certainty. To see this, assume simply that

¹⁹⁵ See chapters III and IV for more on this; the general gist of this would-be objection then is that, due to its ‘radically probabilist’ nature, Jeffrey Conditionalisation ought not to be used in cases where many certainties are involved.

$$P_t(K/H) = P_t(K)$$

i.e. that belief in the Kuhnian meta-induction is independent from belief in any particular theory like H (which seems like a reasonable assumption to make). Further, set

$$P_t(H) = w.$$

Then Jeffrey conditionalising proposition H on K implies that

$$P_{t+1}(H) = P_t(H/K) * t + P_t(H/\neg K) * (1-t).$$

Further, it is the case that

$$P_t(H/K) = \frac{P_t(K/H) \cdot P_t(H)}{P_t(K)} = \frac{r \cdot w}{r} = w$$

and similarly

$$P_t(H/\neg K) = w.$$

All of this together implies therefore that

$$P_{t+1}(H) = w * t + w * (1-t) = w = P_t(H).$$

This implies that reading the uncertainty-inducing arguments has no effect whatsoever on the scientist's belief in the hypothesis in question. This time though, this result is due to the symmetry of the probabilistic independence assumed above, and not to the values of the degrees of belief in any propositions.

What then does this case study show for the above justification for rule BC? Firstly, it points to the consequences of the failure of the rigidity assumption for the workings of the rule. The case study was set up so as to clearly bring this out; however (as was made clear in the previous chapter), this is not the case in real life. Checking whether the rigidity assumption is violated or not is bound to be beyond the capabilities of actual human reasons.

This leads directly to the second lesson of the above case study: if Bayesian Conditionalisation is justified by means of considerations of pragmatic rationality, there are no guarantees that following it will truly have beneficial consequences. The only way to avoid this is by testing the rule, and trying to establish whether any countervailing forces (like the violation of rigidity) are present. This also cannot guarantee success (due to the computational limitations of actual agents), but this is not to be expected from a defence based on pragmatic rationality. A rule that is considered adequate because of its properties of pragmatic rationality can only be fallible and needs to be tested frequently. However, this is less damaging than it might at first appear: there are many situations where the only principles we have at our disposal are fallible.¹⁹⁶ As stated earlier: being fallible is not co-extensive with being useless.

Thirdly, the case study shows that this problem attaches to many other conditionalisation rules that are defended in terms of their pragmatic rationality. In fact, this is simply a *consequence* of such a defence: just like rule BC, Jeffrey Conditionalisation turned out to break down whenever the rigidity assumption is inapplicable.

¹⁹⁶ See also the discussion concerning the philosophy of science in the next chapter.

Finally, it is a natural implication of justifications for the cogency of an updating rule that they do not single it out as central. Once again, though, this is far less damaging than it might at first appear: having a *set* of tools at one's disposal need in no way be inferior to having *one* all-round tool that can be used in all circumstances.

3. Conclusion

I have tried to show that a novel justification for Bayesian Conditionalisation can be based on its pragmatic rationality in the following way: the rule leads to coherently structured thoughts that tend to be useful in helping the agent to avoid acting in a self-defeating way. I have also shown that this novel justification is typical in that it does not single out Bayesian Conditionalisation as the only feasible rule of the probabilistic alteration of beliefs (it is not applicable in the case of the receipt of uncertain evidence for example) and in that it need not always actually produce beneficial results (as evidenced by the case study of the failure of the rigidity assumption). Moreover, there is no principled way of knowing *when* it does so, and when not.

However, these problems were firstly seen not to be unique to rule BC, but as attaching to many conditionalisation rules justified by means of their pragmatic rationality (for example Jeffrey Conditionalisation). Secondly, these problems are all just to be expected from a justification based on pragmatic rationality and are not as damaging as they might at first appear: being fallible and in need of frequent tests does not disqualify a principle from being practically useful.

At this point, the direct discussion of the probabilistic representation and alteration of beliefs has come to its end. What is left to be done is to take stock of what the discussion of the thesis has achieved and what further work could be done in the future. This is the object of the next chapter.

VI. The Probabilistic Representation and Alteration of Beliefs: Summary and Outlook

The present inquiry has now reached its final stage: putting together all the results reached and placing them in a wider framework of further research. I do this in two ways: firstly, by drawing out the implications of the core insights developed in this thesis for more general accounts of the philosophy of science (this will be done in section 1). In a second step, I come to an answer to the question of how compelling the probabilistic representation and alteration of beliefs can be seen to be, particularly with a view to the distinction between logical and pragmatic rationality. I also look out towards further work that could and should be done in this area.

1. The Probabilistic Representation & Alteration of Beliefs and the Philosophy of Science

As argued at various points throughout this thesis, there are some important connections between accounts of the probabilistic representation & alteration of beliefs and the philosophy of science. The aim of this section is to bring these connections out clearly. Note though that I am not defending a particular view of science, but rather seek to illustrate some of the consequences of the inquiry presented here for various discussions in the philosophy of science.

The key aspect in the link between the philosophy of science and the inquiry presented here lies (once again) in the distinction between logical and pragmatic rationality.

Depending on whether one favours a justification of the probabilistic representation and alteration of beliefs based on logical or on pragmatic rationality, one will favour a different conception of the philosophy of science. These conceptions are not necessarily at odds with each other, but do present different outlooks on what scientific activity consists in.

This thesis has presented arguments to the effect that it is logically rational to have beliefs that are probabilistically representable, since it allows an ideal agent to remain faithful to her judgments and preferences. Similarly, it has also been shown that she will use Bayesian Conditionalisation to alter her degrees of belief, as doing so (assuming the applicability of the certainty and rigidity assumptions) is another requirement of this probabilistic representation of beliefs.

This view has a connection to an ‘internalist’ view of science: if science is conceived of as a largely logical venture, where the ideal scientist’s beliefs about various theories stand in a definite relationship to each other, a representation of beliefs based on arguments concerning logical rationality seems very fitting. This view of science might be seen as close to that of the logical positivists, who saw science as progressing in a measured fashion and who believed in logically clearly specifiable relationships between theories.¹⁹⁷

On the other hand, this thesis has also discussed various arguments that tried to defend the probabilistic representation and alteration of beliefs in terms of their pragmatic rationality: it was argued that beliefs which permit of these representations and alterations strongly tend to be helpful to the agent. For example, it was seen as allowing her to avoid

¹⁹⁷ The classic reference here is Carnap 1936/1937. This view of science also fits to that of Howson & Urbach 1993, if they are construed as being directly concerned only with ideal reasoners and with actual reasoners only derivatively. For more on this, see chapter III.

Dutch books in some situations, and generally structured her beliefs so as to help her make decisions that are consistent with each other. This approach to the representation and alteration of beliefs though was seen to be only fallible in practice, without any guarantees as to its actual success in any given circumstance.

This fits to a view of science that is equally fallibilist in nature: on this picture of science, hard-and-fast rules that always give the right result are the exception, not the norm. “Science” is seen instead as a rather messy affair and not a straightforward application of a handful of algorithms.¹⁹⁸ However, fallibility does not mean that such rules should never be trusted; it does mean though that their performance should be regularly checked. It also means one ought to always be aware of the fact that they are merely guiding principles – guiding principles though can still be very helpful devices, particularly in science. Here, it fits well to defend the probabilistic representation and alteration of beliefs in terms of their pragmatic rationality: in a world where hard-and-fast rules are the exception, an equally fallibilist view of the representation and alteration of beliefs seems appropriate.

These two ways of construing science in line with the work of the present thesis can also be used to give an answer to an often-raised problem for the application of probabilistic reasoning to the philosophy of science¹⁹⁹: namely, that it is often seen as very implausible to assume that actual scientists have point-valued beliefs in the truth of the theories they are working with (as seems to be required by probabilistic accounts of

¹⁹⁸ See for example Cartwright 1999, Dupre 1993 and Haack 2003 for more on this.

¹⁹⁹ However, it does not deal with another commonly mentioned problem: that a science based on beliefs is too subjectivist to allow for meaningful scientific disputes. I will not deal with this any further here, though, as it strays too far from the subject matter of this thesis.

the representation and alteration of beliefs). Given the above two ways of construing science, this difficulty can now be answered as follows.

On the one hand, on a picture of science that is based on considerations of logical rationality, the above problem is merely a ‘pseudo-problem’: the beliefs of actual scientists are as irrelevant to this picture of science as the stock of known truths of actual logicians is to deductive logic. This view of science focuses on the relations between theories, rather than on actual scientists’ cognitive attitudes towards these theories. The point-valued beliefs in the above accounts are those of an *ideal* reasoner, not those of actual reasoners.

On the other hand, on a construal of science in line with considerations of pragmatic rationality, it is not even to be *expected* that actual agents can assign point values of degrees of belief to hypotheses of the above sort (see also Howson & Urbach 1993, p. 87, and chapter III of the present thesis). It is rather the case that probabilistic reasoning should be taken as a *qualitative* exercise that has to do with checking the *robustness* of certain hypotheses under changes in their parameters, as well as giving a toolbox of *plausible, qualitative* arguments for ‘good’ scientific methodology (concerning the degree of corroboration of some hypothesis, say). In this way, rule BC is also *not* meant to present a simple algorithm that always gives the right solution for the way in which new evidence that is learned with certainty should be taken into consideration. Instead, its methods should be seen as convenient and helpful ways of making sure one’s inferences are in line with the rest of one’s beliefs, to see *how* dependent they are on these other beliefs and how these beliefs should be sensibly altered in order to avoid making bad decisions concerning one’s research. This explains why actual scientists’ inability to

assign point-values to their degrees of belief is not damaging to this view of the probabilistic representation and alteration of beliefs either: it is simply not part of that view in the first place.

In this way, it becomes clear that the justifications of the probabilistic representation and alteration of beliefs given in the previous thesis are not only worthy of study in their own right, but also have important consequences for the philosophy of science.

2. Final Summary and Outlook

It is now time to briefly recapture where the discussion of this thesis leaves the probabilistic representation and alteration of beliefs, and to look out towards further research that needs to be done in this area.

The first thing to note is that the core distinction running through this thesis is that between pragmatic and logical rationality. The former is to be understood as being concerned with the strong tendency of a certain rule to have beneficial consequences for actual agents, whereas the latter sees these principles as preserving certain assumptions and logical properties and relies on the computational capabilities of an ideal reasoner.

The second point to note is that there are good grounds for seeing beliefs as being representable by means of probabilities, both in terms of logical and pragmatic rationality. Also, it is reasonable to see Bayesian Conditionalisation as a cogent and compelling way of altering beliefs, at least in a number of situations. Moreover, in those cases where the rule is inapplicable (for example in situations based on the receipt of uncertain evidence), other rules like Jeffrey Conditionalisation stand ready to take over.

Of course, this still leaves many questions unanswered. More research is needed in spelling out clearly how accounts of personal identity (and the coherence of a human ‘character’) and the structuring of beliefs due to Bayesian Conditionalisation influence each other. Furthermore, it is crucial to become clearer about how to make sense of our intentional actions (also with a view to the work that is being done in many social sciences and in particular economics), since that was seen to be such a vital part of many of the issues discussed in this thesis. Finally, it is important to find out in more detail how the conclusions of this thesis could be adapted for the *precise* uses of science, particularly with a view to statistical practice. All of this further research is bound to throw up many more interesting insights that will deepen and strengthen the conceptual foundation of our ways of thinking about the world – surely a goal worth striving for.

Bibliography

- Adams, Ernest W. (1964). "On Rational Betting Systems", in: *Archiv fur Mathematische Logik und Grundlagenforschung*, Vol. 6, p. 7-29 / p. 112-128.
- Allais, Maurice (1953). "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axioms de L'Ecole Americaine", in: *Econometrica*, Vol. 21, p. 503-546.
- Bradley, Richard (2005). "Radical Probabilism and Bayesian Conditioning", in: *Philosophy of Science*, Vol. 72, No. 2, p. 342-364.
- Bradley, Richard (2001). "Ramsey and the Measurement of Belief", in: Corfield, David & Williamson, Jon (eds.) (2001). *Foundations of Bayesianism*. Kluwer, Amsterdam, p. 263-290.
- Bradley, Richard (1998). "A Representation Theorem for a Decision Theory with Conditionals", in: *Synthese*, Vol. 116, No. 2, p. 187-229.
- Carnap, Rudolf (1950). *Logical Foundations of Probability*. Chicago University Press, Chicago, IL.
- Carnap, Rudolf (1936/1937). "Testability and Meaning", in: *Philosophy of Science*, Vol. 3, No. 4, p. 419-472 / Vol. 4., No. 1, p. 1-40.
- Cartwright, Nancy (1999). *The Dappled World*. Cambridge University Press, Cambridge.
- Davidson, Donald (1990). "The Structure and Content of Truth", in: *The Journal Of Philosophy*, Vol. 87, No. 6, p. 279-328.
- Davidson, Donald (1980). "A Unified Theory of Thought, Meaning, and Action", in: Davidson, Donald (2004). *Problems of Rationality*. Oxford University Press, Oxford, p. 151-166.

- Davidson, Donald (1976). "Hempel On Explaining Action", in: Davidson, Donald (1985). *Essays on Actions and Events*. Oxford University Press, Oxford, p. 261-276.
- Davidson, Donald (1974). "Belief and the Basis of Meaning", in: Davidson, Donald (1984). *Inquiries into Truth & Interpretation*. Oxford University Press, Oxford, p. 141-154.
- Davidson, Donald (1963). "Actions, Reasons, and Causes", in: Davidson, Donald (1985). *Essays on Actions and Events*. Oxford University Press, Oxford, p. 3-20.
- De Finetti, Bruno (1937). "Foresight: Its Logical Laws, Its Subjective Sources", in: Kyburg Jr., Henry E. & Smokler, Howard E. (1964). *Studies in Subjective Probability*. John Wiley & Sons, New York, NY, p. 93-158.
- Descartes, Rene (1996). *Meditations on First Philosophy: With Selections from the Objections and Replies*, translated and edited by John Cottingham. Cambridge University Press, Cambridge.
- Dougherty, Christopher (2002). *Introduction to Econometrics*. 2nd Edition. Oxford University Press, Oxford.
- Dupre, John (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Harvard University Press, Cambridge, MA.
- Earman, John (1992). *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*. MIT Press, Cambridge, MA.
- Eells, Ellery (1982). *Rational Decision and Causality*. Cambridge University Press, Cambridge.
- Eells, Ellery & Skyrms, Brian (eds.) (1994). *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge University Press. Cambridge.

- Galavotti, Maria Carla (1989). "Anti-Realism in the Philosophy of Probability: Bruno de Finetti's Subjectivism", in: *Erkenntnis*, Vol. 31, p. 239-261.
- Gillies, Donald (2000). *Philosophical Theories of Probability*. Routledge, London.
- Gillies, Donald (1990). "Bayesianism versus Falsifications", in: *Ratio* (New Series), Vol. III, p. 82-98.
- Gillies, Donald (1988). "Induction and Probability", in: Parkinson, G. H. R. (ed.) (1988). *An Encyclopaedia of Philosophy*. Routledge, London, p. 179-204.
- Glymour, Clark (1980). *Theory and Evidence*. Princeton University Press, Princeton, NJ, USA.
- Haack, Susan (2003). *Defending Science – Within Reason: Between Scientism and Cynicism*. Prometheus Books, Amherst, NY, USA.
- Hacking, Ian (1975). *The Emergence of Probability*. Cambridge University Press, Cambridge.
- Hacking, Ian (1967). "Slightly More Realistic Personal Probability", in: *Philosophy of Science*, Vol. 34, p. 311-325.
- Howson, Colin & Urbach, Peter (2004). "The Laws of Probability". Unpublished Manuscript Version of Chapter 3 of: Howson, Colin & Urbach, Peter (2005). *Scientific Reasoning: The Bayesian Approach*. 3rd Edition. Open Court, Chicago, IL.
- Howson, Colin & Urbach, Peter (1993). *Scientific Reasoning: The Bayesian Approach*. 2nd Edition. Open Court, Chicago, IL.
- Hume, David (2000). *An Enquiry Concerning Human Understanding*, in: Beauchamp, Tom L. (ed.) (2000). *An Enquiry Concerning Human Understanding: A Critical Edition*. Oxford University Press, Oxford.

- Jaynes, E. T. (1973). "The Well-Posed Problem", in: *Foundations of Physics*, Vol. 3, No. 4, p. 477-492.
- Jeffrey, Richard (1991). "Introduction: Radical Probabilism", in: Jeffrey, Richard (1992). *Probability and the Art of Judgement*. Cambridge University Press, Cambridge, p. 1-13.
- Jeffrey, Richard (1985). "Probability and the Art of Judgement", in: Jeffrey, Richard (1992). *Probability and the Art of Judgement*. Cambridge University Press, Cambridge, p. 44-76.
- Jeffrey, Richard (1983). *The Logic of Decision*. 2nd Edition. University of Chicago Press, Chicago, IL.
- Jeffrey, Richard (1983b). "Bayesianism with a Human Face", in: Jeffrey, Richard (1992). *Probability and the Art of Judgement*. Cambridge University Press, Cambridge, p. 77-107.
- Jeffrey, Richard (1977). "Mises Redux", in: Jeffrey, Richard (1992). *Probability and the Art of Judgement*. Cambridge University Press, Cambridge, p. 192-202.
- Jeffrey, Richard (1968). "Probable Knowledge", in: Jeffrey, Richard (1992). *Probability and the Art of Judgement*. Cambridge University Press, Cambridge, p. 30-43.
- Jeffrey, Richard (1956). "Valuation and Acceptance of Scientific Hypotheses", in: Jeffrey, Richard (1992). *Probability and the Art of Judgement*. Cambridge University Press, Cambridge, p. 14-29.
- Joyce, James M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge.

- Joyce, James M (1998). "A Non-Pragmatic Vindication of Probabilism", in: *Philosophy of Science*, Vol. 65, No. 4, p. 575-603.
- Kahnemann, Daniel & Tversky, Amos (1972). "Subjective Probability: A Judgment of Representativeness", in: *Cognitive Psychology*, Vol. 3, p. 430-454.
- Keynes, John Maynard (1921). *Treatise on Probability*. Macmillan, London.
- Kolmogorov, Andrei Nikolaevich (1956). *Foundations of the Theory of Probability*. 2nd Edition. Chelsea, New York, NY.
- Lakatos, Imre (1978). "Changes in the Problem of Inductive Logic", in: Worall, John & Currie, Gregory (eds.) (1978). *Mathematics, Science and Epistemology: Philosophical Papers of Imre Lakatos*. Vol. 2. Cambridge University Press, Cambridge.
- Lewis, David (1981). "Causal Decision Theory", in: *Australasian Journal of Philosophy*, Vol. 59, p. 5-30.
- Mas-Collel, Andreu; Whinston, Michael D. and Green, Jerry (1995). *Microeconomic Theory*. Oxford University Press, Oxford.
- Mayo, Deborah G. (1996). *Error and the Growth of Experimental Knowledge*. University of Chicago Press, Chicago, IL.
- McClellan, Edward (2001). "Bayesianism and Independence", in: Corfield, David & Williamson, Jon (eds.) (2001). *Foundations of Bayesianism*. Kluwer, Amsterdam, p. 263-290.
- Mill, John Stuart (1844). "On the Definition of Political Economy", in: Hausman, Dan (1994). *The Philosophy of Economics: An Anthology*. Cambridge University Press, Cambridge, p. 52-69.

- Newbold, Paul; Carlson, William L, and Thorne, Betty M (2003). *Statistics for Business and Economics*. 5th Edition. Prentice Hall, Upper Saddle River, N.J.
- Popper, Karl R. (1999). *Logic of Scientific Discovery*. Routledge, London.
- Ramsey, Frank Plumpton (1926). "Truth and Probability", in: Kyburg Jr., Henry E. & Smokler, Howard E. (1964). *Studies in Subjective Probability*. John Wiley & Sons, New York, NY, p. 61-92.
- Russell, Bertrand (1918). "The Philosophy of Logical Atomism", in: Russell, Bertrand (1986). *The Philosophy of Logical Atomism and other Essays, 1914-1919*. Allen & Unwin, London.
- Savage, Leonard J. (1954). *The Foundations of Statistics*. Dover, New York, NY.
- Schervish, Mark; Seidenfeld, Teddy and Kadane, Jay (1990). "State-Dependent Utility", in: *Journal of the American Statistical Association*, Vol. 85, No. 411, p. 840-847.
- Skyrms, Brian (1986). *Choice and Chance*. 3rd Edition. Wadsworth, Belmont, CA.
- Sobel, J. H. (1988). "Ramsey's Foundations Extended to Desirabilities", in: *Theory and Decision*, Vol. 44, p. 231-278.
- Suppes, Patrick (2002). *Representation and Invariance of Scientific Structures*. CSLI, Stanford, CA.
- Van Fraassen, C. (1984). "Belief and the Will", in: *The Journal of Philosophy*, Vol. 81, No. 5, p. 235-256.
- Von Mises, Richard (1981). *Probability, Statistics and Truth*. Dover, New York, NY.
- Von Neumann, John and Morgenstern, Oscar (1944). *Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, MA.

Williamson, Jon (2006). "Philosophies of Probability: Objective Bayesianism and its Challenges", in: Andrew Irvine (ed.). *Handbook of the Philosophy of Mathematics* (Volume 4 of the *Handbook of the Philosophy of Science*). Elsevier Publishing, Amsterdam.

Williamson, Jon (2005). *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford University Press, Oxford.

Wittgenstein, Ludwig (1921). *Tractatus Logico-Philosophicus*. Kegan & Paul, London.

Worrall, John (1989). "Why Both Popper and Watkins Fail to Solve the Problem of Induction", in: D'Agostino, F. & Jarvie, I.C. (eds.). *Freedom and Rationality: Essays in Honour of John Watkins*. D. Reidel, Dordrecht, Holland., p. 257-296.