The Foundations of Conditional Probability

by

Kenneth Krishnan Easwaran

B.S. (Stanford University) 2002
B.A. (Stanford University) 2002

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Logic and the Methodology of Science

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Associate Professor Branden Fitelson, Chair
Associate Professor John MacFarlane
Professor Paolo Mancosu
Associate Professor Sherrilyn Roush
Assistant Professor Tom Griffiths

Spring 2008

Abstract

The Foundations of Conditional Probability

by

Kenneth Krishnan Easwaran

Doctor of Philosophy in Logic and the Methodology of Science

University of California, Berkeley

Associate Professor Branden Fitelson, Chair

In this dissertation I propose a theory of conditional probability suitable for the interpretation of probability as degree of belief. I first argue that different interpretations of probability must have different mathematical accounts, and then consider both what conditional degree of belief is, and what general form a mathematical theory of it must take. In particular, I suggest that the way to reason about the mathematical formalism is by looking at the uses of conditional probability, in particular in confirmation theory. In Chapter 5, I motivate the central problem that my account of conditional probability addresses, which is that the traditional mathematical account of conditional probability occasionally requires dividing by zero. I then consider alternate approaches to this problem, but reject them because of a series of constraints that I argue any theory of conditional degree of belief must meet. In Chapter 8 I give my positive account, according to which conditional degree of belief must be taken (at least sometimes) to be a relative notion, rather than an absolute one. Where traditional accounts say that it is a function of two propositions, I argue that it depends also on which set of alternatives to the conditioning proposition is relevant. This is a radical proposal, but I argue that in all standard uses of conditional probability, it poses no new problems, because the relevant set of alternatives is specified by the context.

Associate Professor Branden Fitelson
Dissertation Committee Chair

Contents

Acknowledgments

There are too many people that have assisted me in thinking about the topics contained in this dissertation, and in my general intellectual growth through the process of writing it, for me to thank them all individually. This includes not just my past and present professors, colleagues, and students, but also the many people I have discussed my work with from many other institutions. This even includes my family, and friends from outside the department or university, who often put up with me discussing my work at parties and other social events (and sometimes even appeared interested in it).

However, there are a few people I should definitely single out. Mike Titelbaum and Fabrizio Cariani have been two great colleagues to have around for almost my entire graduate career, and all our discussions have been extremely intellectually stimulating. I would also like to thank John MacFarlane and Branden Fitelson for providing so much helpful feedback, both on this dissertation and my other work throughout my years in graduate school. And I am also extremely grateful to Alan Hájek, who has provided feedback throughout this project, and also invited me to spend time with him at the Australian National University, which played a very important role in my intellectual development.

Finally, I would also like to thank the Mellon Foundation and the ACLS, for granting me a dissertation completion fellowship, which allowed me to devote the time and energy to complete this project.

Chapter 1

Introduction

Probability

The mathematical notion of probability has in recent years become important in philosophy, as well as related fields like psychology and economics. In all of these fields, probability is used as a tool to help understand the ways in which people go about forming beliefs, reasoning, acquiring knowledge, and making decisions in the absence of certainty. It might appear that the introduction of a mathematical notion like probability theory to these fields would lead to an increase in rigor, and the ability for greater progress, as in the natural sciences. However, as with all applications of mathematics in fields that are not purely formal, questions arise about both the informal notion to be understood mathematically, and the mathematical formalism that is to be used. If these are not properly dealt with, then the use of mathematics can often confuse things far more than it can clarify them—we would be better off ignoring mathematical methods entirely in favor of the traditional methods of our other fields.

The first type of problem is just a focusing of a problem that already arises before the application of the mathematics. The use of mathematics just makes this problem more pressing, because it can make for more specific disagreements. For the case of probability theory, this is described in philosophy as the problem of the "interpretations of probability". [Hájek, 2007b] Differing answers to this question can apparently give rise to different uses of probability in the gathering of knowledge, as seen by the often heated disputes between so-called "Bayesian" and "frequentist" statisticians, and scientists that argue for the superiority of one or the other methodology. [Jaynes, 2003] Like many philosophers, I will presume that there is no one right answer to this question—many of the proposals that have been made describe (or at least approximate) some useful notion or other for which something like the mathematics of probability is a useful account. The notion I will focus on is that of "degree of belief", which I give a more thorough account of in Chapter 3. However, I do not try here to answer the question that worries statisticians, which appears to be about the question of which notion is appropriate for the best methodology of scientific inference.

Instead, the main part of my dissertation is aimed at the second question—what mathematical theory is the correct one to use? Although [Kolmogorov, 1950] is often cited as the standard exposition of the mathematics of probability theory, there are a variety of other approaches that have been discussed. I don't claim to give a fully-developed mathematical theory of the sort that Kolmogorov did, but just a characterization of what such a theory must look like, and in particular how it must treat the notion of conditional probability.

Even from the beginning, it has been clear that conditional probability poses problems beyond those of unconditional probability. However, on most accounts a theory of conditional probability is necessary to for many of the purposes to which probability theory is applied, such as the understanding of how people's beliefs change over time, and how hypotheses and evidence relate to one another. These applications are central to much of the value of probability theory in its various applications. Thus, a proper understanding of the formal theory of conditional probability is essential to any use of this important mathematical theory, whether in philosophy or any other field.

Miniscule Events

One of the central problems in the formal theory of conditional probability results from the fact that it was originally understood as a ratio: the conditional probability of some event A given B was defined by Kolmogorov to be $P(A|B) = P(A \wedge B)/P(B)$. Because this ratio is undefined when $P(B) = 0$, Kolmogorov was already motivated to provide a second, more complicated, account of conditional probability in the later chapters of his book. The account I defend is a slight variant of his, and I describe it more completely in Chapter 8.

Other approaches to this problem have also been developed. Some [Lewis, 1980] deny that any possibility should ever be assigned probability 0 on the degree of belief interpretation of probability, so that this problem doesn't arise. I rebut the positive arguments for this claim in section 7.1, and I argue in Chapter 5 that in any case, there must be some events whose probability is less than any positive real number, and many applications of conditional probability require conditionalizing on them. I call these events "miniscule" so that I can be neutral about whether they ought to be assigned probability 0 or instead dealt with by some theory using the "infinitesimals" described in [Goldblatt, 1989] and [Robinson, 1996]. But in either case, some new theory must be developed to deal with them.

In addition to the theories involving infinitesimals, another popular approach to this problem of miniscule events gives a new set of axioms on which conditional probability, rather than unconditional, is taken as the fundamental notion. In this way, the problem of miniscule denominators can be avoided, because the ratio account is not used as a definition at all. This approach was first proposed by Karl Popper [Popper, 1959a], but versions of it have also been taken up by others. [Rényi, 1970] Recently, this account has been claimed to be of fundamental importance in our understanding of probability theory. [Hájek, 2003] I argue against this particular account (and others) in Chapter 6, but the positive arguments for my position are given in Chapter 7 and 8.

Relativization

Where Popper's theory leads to the interesting conclusion that all probabilities are really conditional probabilities, mine leads instead to the conclusion that conditional probability must be understood in a relativized way. That is, in general there is no fact of the matter what $P(A|B)$ is, even when we have fixed which person's degrees of belief are being represented by P. Instead, in some cases (in particular, when B is miniscule), this value depends on which set of alternatives to B is under consideration. This may seem to be a problematic conclusion, but I argue throughout Chapter 8 that in the uses we have for conditional probability, this set of alternatives can always be specified by some features of the context. Thus, my account leads us to focus more on this set of alternatives, and the features of the context that determine it, when working with conditional probability.

Outline

Chapters 2 through 5 give the background for my central arguments, clarifying the interpretation of probability I deal with and some general features of this interpretation that distinguish it from others. Chapter 6 considers and rejects alternate approaches to the problem of miniscule events, while Chapters 7 and 8 give my positive account. Finally, Chapter 9 argues that any approach that tries to deny the problem of miniscule events must be committed to a very strong sort of empiricism.

Chapter 2 argues in particular that different interpretations of probability must be dealt with by different mathematical formalisms. Although naive probability theory seems to give a good description of many different phenomena, I claim that this is to some extent a coincidence—the fully worked-out mathematical theories we come up with for different interpretations will be different. In particular, I argue that the notion of conditional probability, which is standardly taken to be a function $P(A|B)$ defined whenever $P(B) \neq 0$, must be defined on a different domain for each interpretation. Thus, even before we consider what values the probabilities must have, we run into differences between the different interpretations.

Chapter 3 gets more precise about the particular interpretation I focus on, of degree of belief. I briefly outline the standard arguments that this notion must satisfy the basic probability axioms, but show there are some problems with these arguments. To avoid these problems, we must avoid any particular operational definition of degree of belief, and must instead just use general coherence between a range of different types of arguments to support the claim that there is in fact a single notion to talk about. I then pursue a similar strategy to understand the notion of conditional degree of belief, and suggest that the way to find out what constraints it must satisfy is to look at the uses we have for this notion—namely, updating, confirmation theory, and decision theory.

In Chapter 4 I start to deal with the mathematical formalism, though I don't yet deal with the actual numerical values that the probability function must take on. I argue that the appropriate mathematical formalism for talking about an agent's degrees of belief must a triple $(\Omega, \mathcal{A}, P)$, where Ω is the space of epistemic possibilities for the agent, $\mathcal{A}$ is some algebra of events, and P is a function defined on $\mathcal{A}$. Leaving out either Ω or $\mathcal{A}$ (as many standard treatments do) misses some features of an agent's doxastic state that are often important. This representation as a triple will play an important role in the later chapters.

Chapter 5 points out that certain aspects of mathematical idealization will not be acceptable. In particular, although agent's minds are finite, this doesn't mean that their doxastic space of possibilities must be finite—in fact, their finitude guarantees that this space must be infinite, in order to accommodate all the possibilities they can't rule out. This means that many events the agent considers epistemically possible must be miniscule. (At this point I don't discuss whether they should have probability 0 or whether infinitesimals are a better approach.) The fact that these events must be conditionalized on in important cases of confirmation theory then raises problems for traditional approaches, as discussed in [Hájek, 2003].

Chapter 6 is fairly short, but is important primarily for the arguments it offers against the use of infinitesimals to deal with the problem of miniscule events. It also discusses alternate approaches to this problem, though it is inconclusive with respect to these others.

Chapters 7 and 8 are the heart of the dissertation. In Chapter 7, I consider several potential desiderata for any account of conditional and unconditional degree of belief, and argue for or against them. In particular, I reject the requirement of regularity, that agents must not assign probability 0 to any event they consider subjectively possible—the fact that a doxastic state includes a set Ω of possibilities eliminates the primary motivation for this requirement, and other motivations are shown to be flawed. I also argue for the notions of countable additivity and conglomerability, as well as two other principles that are essential for my arguments in the next chapter.

Chapter 8 gives the central argument, that conditional probability must be taken to be a relative notion. Instead of being defined as $P(A|B)$, it must be a function $P(A|B, \mathcal{E})$, where $\mathcal{E}$ is a partition of "relevant alternatives" to B. This seems to be a very radical move, but I show that for all the standard applications of conditional probability, this extra relativism is unproblematic, because the relevant partition can be filled in from context. Once I have allowed this extra freedom for the notion of conditional probability to vary based on partitions, and not just the two events, I give some steps towards a method for calculating the values conditional probability must take on. These calculations are fairly tentative, and won't work in every probability space, but they show that there is no need to follow [Hájek, 2003] and take conditional probability as the fundamental notion—conditional and unconditional probability can instead be taken as equally basic notions, (as suggested by [Goosens, 1979]) with constraints relating them, rather than defining one as a special case of the other.

Chapter 2

Subjective Conditional Probability is Different

The bulk of my primary argument is directed at *subjective* conditional probability.[1] Of course, there are also notions of conditional probability associated with epistemic, logical, propensity, frequency, chance, and other interpretations of probability. (See [Hájek, 2007b] for a more in-depth discussion of all of these different interpretations.) At least some of my arguments will make use of premises whose justification depends on the fact that the probability function represents degrees of belief rather than something else. Many of the arguments will most likely have analogues for other interpretations of probability, but I don't expect all of them to.

Some authors have suggested that a theory of probability is inadequate if it doesn't apply to every interpretation of probability. Sometimes this requirement shows up in reverse, suggesting that an interpretation of probability is inadequate if it doesn't satisfy our best formal theory of probability. [Salmon, 1967] At least part of this discussion appears to be merely terminological, so I will largely bracket this sort of issue. Instead, I will throughout focus on providing a good formal theory for degree of belief (and conditional degree of belief) while leaving aside the question of whether this "really" counts as an interpretation of probability.

[1] By "subjective" here I just mean to indicate that the probabilities considered are actual or hypothetical degrees of belief of some real or ideal agent. It may turn out that these degrees of belief are specified in some "objective" way, but I use the word "subjective" to emphasize the fact that these degrees of belief belong to some specific subject. Perhaps some would prefer the use of the term "Bayesian" instead. See Chapter 3 for more details.

However, I will use this chapter to argue that there can't be a unified formal theory that applies to all of the proposed interpretations, in both their conditional and unconditional form. Some may see this as vindicating their claims that one or another of these proposed interpretations isn't really an interpretation of *probability*. However, there are clearly formal similarities between the interpretations—in each case, the unconditional function takes a single input from some algebra of statements or events, and outputs a single real number in accordance with something like the basic Kolmogorov axioms. If this vindicates the claims of others that these really are all interpretations of probability then so be it.

However, I will argue that there must be some differences between the formal systems, so there can't be one mathematical formalism that properly accounts for all of these interpretations. In particular, I will argue that the conditional probability functions must be different. In Chapter 5 I will argue that on the degree of belief interpretation, conditional probability is defined on all pairs of events whose second element is non-empty. In this chapter, I will argue that this is not true on the chance or frequency interpretations, and that a logical interpretation must take a proposition and a *set* of propositions, rather than two propositions. Thus, the different conditional probability functions will have different domains, and thus *a fortiori* will have different formal theories.

Of course, much about one interpretation can be suggestive of related claims about other proposed interpretations, and there are also various principles relating different interpretations of probability. But if the conditional probability functions must have different domains, then it should be no surprise if they take on different values as well. Thus, there may well be different arithmetical constraints relating conditional and unconditional values. These will have to be investigated separately, though of course suggestive analogies can be drawn from work on the other interpretations. Thus, probability theory really is several different theories that happen to share a large core of axioms for some unconditional notions.

But there should be nothing surprising about this. In a sense, what is really surprising is that the formal systems for talking about these different notions are so *similar*. It is quite impressive that there are fairly convincing arguments that each of these notions should satisfy the Kolmogorov axioms, so that they correspond to an additive function from a boolean algebra to the real interval $[0, 1]$.[2] In addition, most of these seem to be fairly

[2]Some approaches may allow for infinitesimals. I argue in section 6.1 that this is probably not the right

plausibly defined over a full σ-algebra (though see Chapter 9 for more on this), and in addition to be countably additive.

Some of this similarity can be explained by the principles (like Lewis' "Principal Principle" and Jaynes' "maximum entropy" principle) relating different probability functions, and other similarities can be explained by analogies between the different phenomena probability is applied to. But it would be a tremendous surprise if these similarities continued all the way to the details of conditional probability on events whose unconditional probability is zero.

2.1 Chance Interpretations

One common interpretation of probability is the "chance" or "propensity" interpretation, proposed by Popper and discussed in [Popper, 1959b]. Several different accounts of this interpretation exist, but the basic idea behind all of them is similar. Some aspects of the world appear to be indeterministic, and in fact according to the standard interpretations of our best physical theories, some of them really are indeterministic.[3] In these theories, there are mathematical quantities attached to indeterministic events that play an important theoretical role. "Chances are posited by scientific theories to explain stable relative frequencies, just as electric charges are posited to explain various observations." [Lange, 2006, p. 399] "Probabilities ... must be physical propensities, abstract relational properties of the physical situation, like Newtonian forces, and 'real', not only in the sense that they could influence the experimental results, but also in the sense that they could, under certain circumstances (coherence), interfere, i.e. interact, with one another."[Popper, 1959b, p. 28] This interpretation focuses on an objective feature of the physical world that seems to obey the probability axioms, and calls this feature "chance". If it turns out that there is no such feature, then this interpretation will turn out to be misguided. But if this feature turns out to have only very slight violations of the probability axioms (for instance, [Lange, 2006] suggests that one might allow for violations of $P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B)$, but only in a few cases, where $P(A)$ is undefined), then it becomes merely a terminological dispute about whether or not this is an interpretation of probability. In what follows, I will assume that there is in fact such an objective feature of the world, and that it basically

way to deal with degree of belief.

[3] Several people have argued that objective chances are in fact compatible with determinism, but see [Schaffer, 2007a] for arguments against this claim.

obeys the probability axioms. But at any rate, as pointed out in [Hájek, 2007b], these "propensities" are themselves unobservable, and we should follow [Lewis, 1970], by seeing what roles propensities play in our theorizing about the world, and letting propensities be whatever things best fill these roles.

The role for the unconditional probability function on this account is to measure the disposition or tendency of a particular experimental set-up to produce a certain frequency of outcomes in the long run—but the question arises whether there is an important role for a *conditional* probability function. Three types of notion seem to be available. One type that some prefer says that *all* propensities are really conditional—any factors relevant to the probability of some event must be explicitly included in the antecedent, so that the propensity for a radioactive particle to decay is given by the probability that it decays, conditional on its being of the relevant isotope, or atomic structure, or whichever feature is relevant. The second type is discussed in [McCurdy, 1996] and [Gillies, 2000], in response to some problems Paul Humphreys raises for conditional propensities. On this account, called the "co-production account" in [Humphreys, 2004], the conditional propensity $P(A|B)$ is just the propensity of the relevant system to produce both A and B, divided by the propensity to produce B. (Nothing is said in either discussion about how to interpret this when $P(B) = 0$—Gillies mentions this problem briefly on p. 831, only to explicitly bracket it.) But perhaps the most obvious notion is one suggested in [Humphreys, 1985]—if an unconditional propensity is some sort of unconditional disposition (for instance, the disposition to decay that makes a particle radioactive gives its probability of decay), then a conditional propensity should be some sort of conditional disposition (for instance, the disposition to break if hit that makes a vase fragile gives the conditional probability of its breaking, given that it is hit). At any rate, each account comes in many varieties, depending on whether the relevant propensities are propensities of situations to generate specific single events, or propensities of repeatable sets of conditions to produce certain frequencies in the long run, or something else.

The co-production and conditional disposition types of account are similar in that both index the propensity to a time and situation. Some authors indicate this by adding an subscript to the "P" for the time, and possibly the world as well, while others instead explicitly include all the background conditions and history of the world up to that time among the propositions conditioned upon. (Humphreys and McCurdy, confusingly, seem to use both notations in every formula, giving things like "$Pr_{t_1}(T_{t_3}|I_{t_2}B_{t_1})$" to indicate the probability, at t_1, with background conditions B_{t_1}, of T occurring at t_3, given that I occurs at t_2. This seems like strange notation, because I and B seem to play very different roles in the probability statement—B gives the background conditions with which all the probabilities are assessed, while I is just some specific event that may or may not occur in the future.) The first account differs by requiring every statement of propensity to be conditional on sufficient information to give well-defined propensities for various events. In some sense, this is the most natural way to understand propensities when they are thought of as a sort of modification of a frequency account of probability. [Popper, 1959b, p. 34] However, as suggested by the accounts discussed in [Gillies, 2000, pp. 822-825], it's not always clear that the background information B can be given in any sort of propositional form—it may be best understood as a physical situation or something else of very different form than the events whose probability we are interested in. But even if it *is* understood in the same way as them, then we certainly do not have a system where $P(A|B)$ exists iff $P(B|A)$ exists. If B is something like "a standard American penny is flipped fairly" and A is "it comes up heads", then it looks like $P(A|B)$ makes sense (and in fact is very close to 1/2), but $P(B|A)$ doesn't—specifying that a coin comes up heads doesn't do enough to properly characterize a situation that has a stable propensity to involve a standard American penny being flipped fairly. However, if P is meant to be the credence function for a particular agent at a particular time, then it looks like $P(A|B)$ and $P(B|A)$ should equally make sense. Thus, this first account of conditional propensities looks like it must involve a different sort of function than the one for conditional subjective probabilities, so there is no reason to assume these different "interpretations of probability" obey the same mathematical formalism.

However, there seem to be some good reasons to prefer the other two accounts of conditional propensity over this account, on which *all* propensities are conditional. McCurdy motivates this by pointing out that there seems to be a distinction between the role of "background" propositions specifying the conditions in effect at the assessment of the probability function, and the propositions that can have probabilities and can be conditioned on.[4] He calls the set of possibilities relevant for the latter "the event space", and says that "an asymmetrical relationship between the background conditions and the event

[4]Gillies also makes this distinction, but considers both sorts of events as entering into conditional probabilities. He just calls probabilities conditional on the background "fundamental conditional probabilities", and ones conditional on other propositions "event conditional probabilities." [Gillies, 2000, p. 828]

space is created, and this relationship is not formally represented in other interpretations of the probability calculus." [McCurdy, 1996, p. 109] I agree that this asymmetrical relationship should hold for propensities, but I think it should also hold for degrees of belief. (See Chapters 3 and 4.) However, this similarity is not sufficient to ensure that the formal requirements for conditional probability will be the same for propensities and degrees of belief.

The co-production interpretations of conditional propensity give rise to some special problems. First of all, it seems in some sense that these are just definitions of conditional propensity that are designed to suit the mathematical axioms, rather than having some specific interest of their own. If such a function can be useful in calculations, that's fine, but it doesn't give rise to an interesting notion whose analysis can be given correctly or incorrectly, the way the other two types of account do (and the way I think degree of belief does—see Section 3.2). But worse, there seem to be important uses for conditional propensities that the co-production account gets wrong. As Humphreys points out, "at best it preserves probability theory at the expense of losing the characteristic dispositional content of conditional propensities." [Humphreys, 2004, p. 673]

Consider a fair coin-flipping machine whose trigger is inside a locked room. The easiest way to open the door is to turn on a huge magnet, which unlocks the door, but also prevents the coin from rotating in the air, so that while the magnet is activated, the coin almost always lands heads. Then on the co-production account, the probability of the coin coming up heads given that it is flipped is close to 1, because the propensity of the situation to give rise to a coin landing heads is almost as great as the propensity of the situation to give rise to a coin being flipped at all. However, while the magnet is off and the room is locked, it seems intuitively that we want to say that the propensity for the device to produce a coin landing heads, given that it produces a flip at all, is close to 1/2—this is exactly what we mean to say that it is a fair coin-flipping machine. This example shows that the co-production account of conditional propensities gives rise to a notion that is not actually dispositional—it is a parallel to examples of finks that are standard in the literature on dispositions (see for instance [Fara, 2006]).

It might be objected to this apparent counterexample that often this sort of non-dispositional behavior is exactly what we want for conditional propensities. For example, [Gillies, 2000, p. 828] includes an example of a barometer, in which it seems that we want a notion of conditional propensity for which the probability of a storm, given that the barometer drops, is fairly high, even though there is no causal influence of the barometer on the storm. Now it may be that we want such a notion of conditional *probability*, but it's not clear that a notion of conditional *propensity* is needed here. I think all the work of this conditional probability can be done with a notion like conditional *degree of belief*, even though an important part of the work here will be done by the notion of *unconditional* propensity (or objective chance).

To see how this works, we can observe how the expected conditional probability comes out of the unconditional chances by means of Lewis' "Principal Principle." [Lewis, 1980] A simplified version of this principle states that if an agent knows at some time t the chance at that time that some event A will occur, then her degree of belief in that event should equal the chance of that event.[5] Now if the agent knows early on the chances of storms, low-pressure systems, and drops in barometers, and the correlations among these three types of events, then the Principal Principle means that her credences will also match these correlations. Then the expected conditional probabilities show up as the agent's conditional credences, so that if drops in the barometer and storms are both strongly correlated with low-pressure systems, then conditioning on one will give a higher credence in the other. In more realistic cases, the agent doesn't know all the chances of these three types of events, but she does know at least some of the correlations (or at very least, she has a high degree of belief in the correlation between drops in the barometer and storms)—working through Lewis' account of propositions conditional on statements about chances, one can verify that the beliefs about correlations show up as correlations in the probability function representing the agent's degrees of belief.

Given that an account of conditional chance doesn't need to accommodate these non-causal correlations, it seems that we can get a more useful and interesting theory by taking the third option—just as chances are dispositions on the propensity account, conditional chances are conditional dispositions. As Humphreys puts it in his reply to Gillies and McCurdy, "the conditional propensity constitutes an objective relationship between two events and any increase in our information about one when we learn of the other is a completely separate matter." [Humphreys, 1985, p. 563] But this ends up leading the conclusion I mentioned above—conditional propensity can't be understood with the

[5]Lewis formulates the principle in terms of an initial probability function conditional on all evidence that the agent knows, in situations where the agent knows no "inadmissible evidence". In later chapters I explain why I think using a single initial probability function is not the right way to go. The details of what makes evidence admissible or not are inessential here.

same mathematical formalism as conditional degree of belief (or indeed with the standard framework for conditional probability).

To make the minimum modification to the standard framework, one natural suggestion (embraced by Fetzer and others) is that in cases like the ones Humphreys describes, conditional propensities are defined in one direction, but undefined in the other. That is, $P(A|B)$ makes sense when A is the event of a coin flip coming up heads, and B is the event of it being flipped fairly, but not when the two are reversed. (This is a difference from Humphreys' picture, on which the inverse conditional propensity has a value of 0, 1, or the value of a related unconditional probability, rather than being undefined.) Then, we just require that conditional propensities obey the standard probability axioms whenever they are defined.

However, the example of the coin flipping device in a magnetic room seems to suggest otherwise—in that case, the only way to get the intuitively correct conditional probabilities involves violating the ratio account of conditional probability, even though there is no division by zero or anything else mathematically strange going on. But I think that making this much of a departure from the standard Kolmogorov axioms for probability makes sense in this case, at least in part because of the metaphysics of the objects of the probability functions.

Humphreys suggests that "the arguments of the propensity functions are names designating specific physical events. They do not pick out subsets of an outcome space as in the measure-theoretic approach." [Humphreys, 2004, p. 669] This, if correct, would be a further distinction between degrees of belief and propensities. Although it doesn't immediately suggest mathematical differences between the two functions, it can help suggest why the conditional probabilities don't always follow the ratio—this is because events are not subsets of each other on this picture, the way they are on the set-based picture due to Kolmogorov (which I argue in Chapter 4 is correct for degree of belief).

Further considerations on conditional chance come up when considering its uses. Humphreys' paradox (which motivates moving away from the standard mathematical formalism) proceeds from an intuition about what conditional chance means. One might suspect that other uses of conditional chance push towards the more standard account. In the barometer case mentioned above, this appears to be true, but it seems that using Lewis' Principal Principle, all the work that conditional chance seems to do there can be done by conditional degree of belief. In [Lewis, 1980], Lewis suggests another use for conditional chance, which is to describe how *unconditional* chances change over time.

The idea is that (unlike in the account of conditional chance that makes all chances conditional), chances are explicitly indexed to times. Thus, at two different times t_1 and t_2, there are generally two distinct probability functions P_1 and P_2 giving the chances for various events. Lewis argues that $P_2(A) = P_1(A|B)$, where B is "the complete history of the interval between" the two times. [Lewis, 1980, p. 280] In [Lange, 2006], there are some arguments against this claim—however, these depend on whether objective chances can change without any particular non-dispositional fact being the basis of such a change. At any rate, the modified claim is that instead of B being the history of all events occurring in the interval, B is specified to be the conjunction of outcomes of all chance events in the interval.

But in either case, only very specific conditional probabilities are needed—these are always probabilities of a later event conditional on an earlier one (so none of Humphreys' inverse conditional probabilities need to be used), and these are in fact always probabilities conditional on complete sets of occurrences between two times. It is plausible that such cases are never like the case I described with the coin-flipping device in a magnetized room, so that the question of whether those sorts of cases violate the ratio account of conditional probability also doesn't come up. Thus, this use of conditional chance doesn't seem to cut one way or the other in a dispute about whether conditional chances should behave like other conditional probabilities. If anything, this use suggests that Humphreys may be right about the order of the events being relevant to whether the conditional chance is even defined.

2.2 Logical Interpretations

The existence of a logical interpretation of probability is widely contested. The notion is most closely associated with [Carnap, 1950], but it has been discussed by many others. The basic idea is that this notion is supposed to provide some sort of generalization of the notion of deductive entailment, and can hopefully make sense of the notion of confirmation that scientists use when they suppose that some evidence supports one theory or another.

A fundamental fact about this notion of probability seems to be that it is inherently a *conditional* probability notion, and not an unconditional one. This claim was supported both by classical endorsers of the theory, like Carnap, but also contemporary ones like Patrick Maher. "It is clear that this concept of confirmation is a relation between two sentences, not a property of one of them. Customary formulations which mention only the hypothesis are obviously elliptical; the basis is tacitly understood." [Carnap, 1945, p. 515] "p is a function that takes two sentences (or, alternatively, propositions) as its arguments, has real numbers as its values, and satisfies some version of the mathematical laws of conditional probability. ... The function p is to be defined by specifying those σ_1 and σ_2 for which $p(\sigma_1, \sigma_2)$ exists and giving rules that logically determine the numeric value $p(\sigma_1, \sigma_2)$ for all such σ_1 and σ_2." [Maher, 2006b, p. 518] Thus, the suggestion seems to be that the conditional probability notion is fundamental for logical probabilities. As I argue in section 6.2, this is not true for subjective probabilities, and we have seen already that it is also false for chance or propensity interpretations of probability. Thus, there is already a distinction in the formal mathematical notion for these distinct interpretations of probability.

However, I think that Maher doesn't go far enough. When he states that inductive probabilities are relative to evidence, he gives no argument that evidence is best explicated as a single proposition, rather than as a body of propositions. Whenever the body of propositions is finite, it may be possible to replace it by the conjunction of the propositions it contains. Some infinite bodies of propositions also have a single proposition that is logically equivalent to their conjunction. However, it's not clear why the body of propositions should always be either finite, or one of the special infinite sets that has a finite conjunction. And in fact, elsewhere he does take this step: "In any case, the interpretation of joint confirmation as confirmation by a conjunction strikes me as unnatural. What I propose instead is that we think of confirmation in the general case as a relation between a *set* of pieces of evidence and a hypothesis." [Maher, 1996, p. 163, emphasis in original] If true, this would be a more radical break with other interpretations of probability. On other interpretations, when there is a conditional probability notion, it is a function of two propositions. If one proposition must be understood instead as a *set* of propositions, then standard formalisms will not work.

First-order deductive logic has a compactness theorem stating that if Γ is an infinite set of sentences that entails ϕ, then there is a finite set $\Gamma_0 \subset \Gamma$ which itself entails ϕ. Since we can take replace Γ_0 by the conjunction of its finitely many members, we can thus show that in some sense all entailment in first-order deductive logic in some sense derives from an entailment relation between two sentences—there is a sense in which the general relation between a sentence and a set of sentences can be thought of as derivative from this other notion.

We might hope that something similar could be done for *inductive* logic too. However, there is a significant problem for anything like the compactness theorem holding in the inductive case, which is that inductive logic is non-monotonic. Deductive logic is monotonic, meaning that if a set of premises entails a conclusion, than adding extra premises cannot destroy the entailment. However, with inductive logic, it is clear that however high or low the support a set of premises gives a conclusion, provided it is short of deductive entailment, there are further premises that can be added to drastically change the degree of support. Thus, there is no hope of representing the relation between an infinite set of premises and a conclusion by means of some finite subset of the premises, because we would need a further guarantee that the infinitely many premises that are left out are for some reason irrelevant.

Regardless of this special problem for inductive logic, there is a more general problem with replacing any logical notion of entailment or support by a notion that relates two sentences, rather than relating a sentence to a set of sentences. This problem is that there are many non-compact logics. For instance, although first-order deductive logic is compact, *second*-order deductive logic is non-compact. If P is some predicate, then the second-order Peano axioms together with $P0$, $P1$, $P2$, etc. entail $\forall x Px$, but no finite subset of those sentences does. Thus, if logical entailment is allowed to include second-order logical relations (or those of any other non-compact logic) then there is no hope of replacing the standard entailment relation by a relation between two sentences. Thus, even if Maher's particular notion of logical probability is not the right one, it seems that any such theory should take the relevant notion of probability to fundamentally be a function from a proposition and a set of propositions to a real number, rather than a function on a single proposition (as probability is standardly understood) or a function on a pair of propositions (as conditional probability is standardly understood).

2.3 Frequency Interpretations

Much has been said for and against frequency interpretations of probability. They are normally seen as competitors to the chance or propensity view (at least, for making sense of the probabilities that arise in physics or the other sciences), though of course even if the propensity view makes sense and exhausts our scientific uses of probability theory, frequencies certainly give a function that can be usefully called a probability. It's well-known [van Fraassen, 1977] that limiting relative frequencies violate many of the formal criteria that other interpretations of probability do. For instance, they aren't countably additive, and the domain they are defined on isn't closed under either countable or finite intersections. The additional point that I will make is that conditional probabilities often fail to be defined on this account, even when both events have well-defined non-zero probabilities. The example is just a slight modification of the example van Fraassen gives to show that the domain of the relative frequency function is not closed under intersections.

Consider a countably infinite sequence of trials, as is the standard background for making sense of relative frequencies. Let X_1 be the first two trials. For greater n, let X_n be the set of all trials after the 2^{n-1}th up to (and including) the 2^nth. The event A, which consists of the union of the X_n for odd n, has no limiting relative frequency, because its relative frequency fluctuates from 1/3 to 2/3 infinitely often. Let B be the event that occurs exactly on all even trials within A, and all odd trials outside A. Let C be the event that occurs on all even trials, whether in A or outside it. As van Fraassen notes, $B \cap C$ has no limiting relative frequency, because when n is even, the number of occurrences of $B \cap C$ up to n is exactly half that of A, so its relative frequency fluctuates from 1/6 to 1/3 infinitely often. But this also means that the conditional probability $P(B|C)$ must be undefined (at least, on standard understandings of conditional relative frequency). The ratio account of conditional probability clearly leaves it undefined, because the numerator, $P(B \cap C)$, is undefined. But even without this definition, if we consider the number of occurrences of B in the first n occurrences of C, we will find that this relative frequency also fluctuates infinitely often between 1/3 and 2/3, so that this conditional probability is undefined.

As I will argue later, conditional subjective probability is defined whenever the event conditioned on is not an epistemic impossibility for the agent. Conditional chances sometimes fail to be defined (because the causal relations between the events are backwards) and sometimes fail to conform to the ratio when it is defined. Logical probabilities are inherently conditional, but they conditionalize on sets of propositions, rather than individual propositions. And finally, conditional frequencies sometimes fail to be defined even when both propositions have well-defined frequencies, but this failure is of a different sort than the failure of conditional chances to be defined. Thus, each of these accounts must give rise to a different notion of conditional probability, and thus each should be investigated separately, though they may still shed some light on one another.

Chapter 3

Subjective Probability

3.1 What is subjective probability?

First of all, when discussing subjective probability, I don't mean to imply that any of our ordinary uses of the word "probability" refer to this notion—arguments against such a claim are given in [Maher, 2006a, pp. 187-9]. However, two other claims are implicit in discussing a notion of subjective probability. One is that there are states one can be in that are properly called degrees of belief (or credences, or partial beliefs, or something similar). I take it that some aspect of this claim is uncontroversial. The second claim, known as "probabilism", is a bit more controversial, stating that these degrees of belief obey some standard axiomatization of probability theory. Some people have been skeptical of these claims—for instance, consider [Kyburg, 1978, p. 179]: "The theory of subjective probability is psychologically false, decision-theoretically vacuous, and philosophically bankrupt." But as far as I can tell, the arguments provided there are all inconclusive, and many of them presuppose a particular reduction of degree of belief to betting ratios.

This reduction goes back at least to [Ramsey, 1926]. The idea is that ideally rational agents always have a fair betting price for any proposition A such that the agent is willing to either buy or sell for $\$Pr(A)$ a gamble that pays \$1 if A is true and nothing if A is false. It can then be shown that if the function Pr doesn't satisfy the Kolmogorov axioms of probability theory, then the agent is susceptible to a so-called "Dutch book", which means a series of bets that she is willing to accept individually, which collectively guarantee that she will lose money. Thus, Ramsey identified subjective probability with this function Pr. There are various problems with this picture, but Kyburg is too fast in assuming that these problems doom the theory of subjective probability as a whole.

An older tradition, going back at least to Bayes and LaPlace suggests that degree of belief is somehow connected to decision making (whether in the simple gambles Ramsey and others considered, or the more complicated gambles considered by decision theory, as described by Daniel Bernoulli and others). Ramsey's project can be seen as a positivist or behaviorist extension of this tradition. Rather than saying that degree of belief is *connected* to this behavior, he *identified* it with the behavior.

Now as a matter of fact, it is clear that most people do *not* have such a fair price for most propositions, such that they're willing to accept bets for or against this proposition at that price. In addition, there may be a case to be made that a certain amount of risk aversion is rational even under idealized circumstances, so that one can't even argue that *idealized* agents have these fair prices. However, such agents are not omniscient, so they would still have intermediate degrees of belief, and thus their degrees of belief *can't* just be the betting odds they would accept.

Given that we have abandoned positivism and behaviorism, there is no need to identify subjective probability with something purely operational like betting behavior. Instead, we can follow [Christensen, 1996, pp. 457], and say that "degrees of belief may . . . *sanction* certain betting odds, even if the degrees of belief do not *consist in* propensities to bet at those odds". If we assume that for every bet, an agent's degrees of belief sanction either that bet or its opposite, then we can run a version of the Dutch book argument to show that these degrees of belief must correspond to a probability function. That is, if an agent's degrees of belief fail to satisfy the axioms of probability, then there is a set of bets such that the degrees of belief sanction them individually as fair, and yet together they lead to a sure loss for the agent. Therefore, since rational degrees of belief shouldn't sanction as fair a set of bets that logically guarantees a loss for the agent, there is still an argument that degrees of belief should satisfy they probability axioms, even after giving up their identification with betting behavior. Following [Eriksson and Hájek, 2007], one can use arguments like Kyburg's to contest the adequacy of *any* such operational definition of degree of belief, while still allowing that an agent's betting behavior is a very reliable indicator of her degrees of belief. The fact that the theory of subjective probability has led to a large and flourishing body of work suggests that more conclusive arguments than Kyburg's are needed to abandon the project of probabilism.

One might worry that although the idea of degrees of belief has been quite useful, the appropriate formalism for discussing it is not probability theory. But there are other arguments for probabilism. One popular type of argument is based on decision theory—since the middle of the 20th century, it has been clear that an agent's preferences among actions can satisfy a certain set of axioms iff they can be represented by a credence and a utility function. The utility function must give the agent's preferences among known outcomes, and the credence function must satisfy the axioms of probability theory. Versions of this argument are discussed extensively in [Joyce, 1999]. Like the Dutch book arguments, these arguments are problematic because they essentially invoke pragmatic factors in justifying apparently epistemic principles. Additionally, there are further worries about what connection the utility and credence functions have to the ordinary notions of desire and belief that proponents claim they represent. However, these arguments do give a kind of independent support for probabilism.

Yet another independent argument for probabilism dates back to [Cox, 1946], and is used as a foundation by [Jaynes, 2003] and others for their more objective version of Bayesianism. This argument starts from some very plausible premises, and argues that degree of belief should be represented by a function B satisfying certain conditions. The premises informally state that rational degrees of belief can be represented by real numbers, that they have certain connections to one another, and that they are in a certain sense unique. In particular, the degree of belief in a proposition is sufficient to determine the degree of belief in its negation, so there is a function f where $f(B(p|q)) = B(\neg p|q)$; and the degree of belief in a conjunction is also determined by the (conditional) degrees of belief in the conjuncts, so there is a function g where $B(p \wedge q|r) = g(B(p|r), B(q|p \wedge r))$. From the additional assumptions that f and g are both differentiable, and that $B(p|q) = 1$ when p is a tautology, and $B(p|q) = 0$ when p is a contradiction and q is not, one can prove that there is some real number x such that $(B(p|q))^x$ satisfies the standard Kolmogorov axioms. (These premises also end up supporting a particular version of the "Principle of Indifference", which I will discuss a restricted version of in section 7.4.) However, these principles required are fairly strong, and it's not totally clear why they must be obeyed by rational agents (see [Colyvan, 2004] and references therein).

Although I have given some indications of some problems with all of these arguments, there are still others available. [Joyce, 1998] Thus, I agree with Christensen:

> Neither [the Dutch book nor the Representation Theorem] comes close to being a knock-down argument for Probabilism, and non-probabilists will find con-

> testable assumptions in both. But each one, I think, provides Probabilism with interesting and non-question-begging intuitive support. And that may be the best one can hope for, in thinking about our most basic principles of rationality. [Christensen, 2001, p. 375]

3.1.1 Descriptive vs. Normative

A common objection to probabilism as a theory of degrees of belief is that people in fact often fail to obey the axioms of probability theory, for instance in the famous "conjunction fallacy" of [Tversky and Kahneman, 1983]. However, this is only an objection to probabilism if it is taken as a descriptively accurate theory of how people actually reason. At least in philosophy, probabilism is taken to be a normative theory. But there is an apparent problem to this interpretation as well—it is an empirical question whether human beings are constitutionally capable of forming doxastic attitudes other than certainty in a proposition, certainty in its negation, and withholding belief. If not, then a normative theory of subjective probability can't give norms of rationality for humans, and so it would be just about useless. (It could still have a role in the design of artificial intelligences and the like, but this is hardly the sort of role normative theories in philosophy tend to aim at.) Therefore, there seem to be problems with a philosophical theory of subjective probability, whether descriptive or normative.

However, there is room between the horns of this dilemma—as long as the empirical question of whether humans can have intermediate degrees of belief is answered in the affirmative, then there is room for a normative theory about how these degrees of belief ought to be arranged. Philosophical work on subjective probability is therefore only relevant to the extent that humans (or rational agents in general) are capable of having degrees of belief.

Fortunately, psychologists have made empirical arguments suggesting that something very much like degree of belief is in fact a part of human mental life. [Tenenbaum and Griffiths, 2001] In addition, they have suggested that degree of belief as a descriptive matter of fact does tend to conform closely to the axioms of probability, at least in many ordinary circumstances, when the hypotheses under consideration are properly identified. However, this extent of descriptive accuracy of the theory is conceptually unnecessary to defend philosophical subjective probability theory—all that is important is that there actually be a notion of degree of belief that is appropriately connected to other notions (like fair bets, and as I will discuss later, confirmation), and then the arguments will be properly grounded. The fact that the normative aspects of the theory are at least in part descriptively accurate gives a certain confirmation to the idea that we have found the right norms, since we expect people to generally function quite well as rational agents. But there is no reason for us to assume that humans function perfectly, so it is fine for the theory if people as an empirical matter of fact do violate it from time to time. Thus, the theory of subjective probability has a descriptive component, but can be largely pursued as a normative theory once that component has been demonstrated. Further empirical work can be helpful (especially if it suggests that humans may massively violate the normative theory, as in the work of Kahneman and Tversky) but does not have to be central.

3.1.2 Subjective vs. Objective

Among the theorists that talk about probability as some sort of notion of "rational degree of belief", a distinction is sometimes drawn between "subjective Bayesians", who allegedly say that any function consistent with probability theory is rationally permissible for one's degrees of belief, and "objective Bayesians", who allegedly say that there is in fact a unique function that every rational agent should have. My position is intermediate between these (see Chapter 7 for many principles that I endorse that go beyond the basic axioms of probability theory), and I suspect that the same is true for the majority of theorists. Whether this makes my position "objective" or "subjective" in the sense of this distinction is just a terminological question. In addition, it is also irrelevant for my central claims whether further principles are also true, that would take the requirement much closer to the objective end. The ones I endorse in Chapter 7 are the only ones that are needed for my arguments.

However, there are some proposals that look much like objective Bayesianism that are really something different. In particular, I am thinking of the "evidential probability" discussed in [Williamson, 2002, Ch. 10]. Williamson explicitly argues that these evidential probabilities may not be the credences of any agent, even a hypothetical perfectly rational agent. Instead, he says they are probabilities given by the evidence. In the rest of the chapter he argues that these probabilities play an important role in epistemology, and his constant comparisons to Bayesianism might lead the reader to believe that he is proposing a different theory of something like rational degree of belief.

But in other respects, this probability function really is more like a logical probability function (as I discuss in section 2.2) than like the degree of belief functions that I'm primarily interested in. In particular, he admits that an agent may know a proposition, and therefore have it as part of her evidence, so that its evidential probability is 1, and yet the agent is not subjectively certain enough in it to have degree of belief equal to 1. In fact, he argues that this situation is often rational. Thus, evidential probabilities are not degrees of belief. Instead, they give the relation between pieces of evidence and a potential conclusion.

However, unlike Maher, Williamson seems to treat the unconditional probability notion as basic here, rather than the conditional one. "$P(h|e) = P(h \wedge e)/P(e)$; this equation only makes sense if e is propositional. We therefore cannot ... [deny] that evidence is propositional, for then evidential probabilities would be undefined." [Williamson, 2002, p. 213] However, this argument has force only if we adopt the ratio account of conditional probability. If (as I suggest in section 2.2) the best way to understand a body of evidence is as a *set* of propositions, rather than just a single proposition, then this argument must fail. Williamson will need to adopt a theory of evidential probability on which conditional probability is a primitive, rather than being definable from unconditional probability by the ratio formula.

Even so, there are clear connections between this evidential probability (which is not my primary subject in the later chapters) and something like an objective Bayesian notion of degree of belief (which I do take to be within the scope of my later discussion). On pp. 222-3, Williamson argues that there is in fact a norm connecting evidential probability to degree of belief. "Proportion your belief in a proposition to its probability on your evidence." He goes on to argue that this rule is a good one, even though we can't always know what our evidence is, and thus can't always know whether we are following the rule. Maybe this rule is generally a useful one, but I claim that it can't be quite as strong as one might suppose. If epistemic probabilities obey principles that degrees of belief must violate, then one can't in general make the two line up. Williamson accepts this fact, because of the general non-transparency of when one is following any rule with non-trivial conditions. Thus, this rule may well be compatible with the rest of the norms that I propose, even if the norms end up requiring that evidential probabilities and degrees of belief sometimes come apart.

3.2 What is *conditional* subjective probability?

In the initial Kolmogorov axiomatization of probability, unconditional probability is taken as the basic notion, and conditional probability is taken as a subsidiary one mathematically defined in terms of it. However, thought of this way, there is nothing conditional about conditional probability—it is just one mathematical function of two variables, defined in terms of a function of one variable. This function is sometimes interesting for technical mathematical reasons, as in the "probabilistic method" in combinatorics. However, the rest of the time, I suspect that this abstract function is of interest precisely because it comes close to capturing some intuitive notion. This notion is what I am interested in, and it must also be the target of any attempt to found probability theory on conditional (rather than unconditional) probability. However, just what conditional probability is is even more unclear than for unconditional probability.

Several candidate approaches to conditional probability are the following: an update plan, the probability of a conditional, and "the Ramsey test".

On the updating account, $P(A|B)$ represents the degree of belief an agent *would* have (or ought to have) if she were to learn (with certainty) that B is in fact the case. Most philosophers discussing this matter have suggested that in ordinary cases, an agent ought to update by moving to a credence function that agrees with the ratio account, so this would explain the connection between the formal mathematical notion and the conditional notion here. (See [Greaves and Wallace, 2006] for one such argument.) However, it's not clear that this notion really defines a two place function whose arguments are both of the same type as the unconditional probability function. Even in ordinary cases, it takes a lot of idealization to claim that there is a single proposition that an agent learns between one time and another. In many cases, it seems there are infinitely many such propositions, and it's not clear that an agent's algebra of events will always be closed under such infinite conjunctions—thus, this conditional notion would have more in common with those for logical or "epistemic" probabilities, mentioned above. In other cases (famously discussed by Richard Jeffrey), it seems plausible that there is *nothing* that is learned with certainty, and yet the agent's degrees of belief change rationally. Finally, there are also many cases in which an agent's degrees of belief change by *loss* of certainty, rather than gaining new knowledge. Thus, if update is the defining feature of conditional probability, it would seem strange to focus on the case of gaining a single proposition with certainty, rather than including all the cases of loss, multiple propositions, and less-than-certainty. It seems clear that conditional probability is connected to update somehow, but it is implausible that it is *constituted* by update.[1]

Another possibility one might consider, inspired by Ernest Adams and others, is to say that the notion of conditional probability just is the degree of belief assigned unconditionally to an indicative conditional. Unfortunately, a series of triviality results (see for instance the papers in [Eells and Skyrms, 1994]) have shown that such an identification would lead to triviality of the probability function, assuming some standard connections between conditional and unconditional probability. In addition, this analysis presupposes that an indicative conditional really is the sort of thing that can be assigned a degree of belief, which is a major debate in semantics. At any rate, even if this identification could be made to work, it seems that the reduction would go the other direction—indicative conditionals would be understood in terms of conditional probability, rather than the other way around. Thus, it will be no help in analyzing conditional probability.

Ramsey Tests

A more careful suggestion is given by what [Bennett, 2003] calls the Ramsey test—the conditional probability $P(A|B)$ is given by the degree of belief one has in A when supposing B, or "hypothetically adding it to one's stock of beliefs". This may often differ from how one's beliefs would *actually* change were one to learn B with certainty (even idealizing away learning anything else). That is, it might be a plan for updating, without being how one actually will update. (Some of the problems with actually following through on these plans are brought up in [Gaifman, 2004, p. 115].) Even if it is the plan one *ought* to have for updating, it doesn't follow that when the time comes one *ought* to follow it. The Ramsey test may also give a different value from one's belief in a conditional, because of the triviality results and worries about semantics of conditionals. However, like conditionals but unlike updates, supposition generally features a single event (or a finite conjunction of them) rather than an infinite set of premises. Additionally, worries about lack of certainty or the loss of information are irrelevant—there is no such thing as "partially supposing" a

[1] A further point (for which I thank Mike Titelbaum) is that this would in some sense compress two norms into one—if conditional probability is a mental state of an agent (as the subjective account of probability would seem to require), then the norm connecting it to unconditional probability is synchronic, while the one connecting it to update is diachronic. This update account prevents the isolation of these two separate norms.

proposition, or "negatively supposing" something.

Note that some care must be taken with the notion of supposition. This problem is brought out quite pointedly in [Chalmers and Hájek, 2007], where they point out that on a certain notion of the Ramsey test, together with some reasoning based on Moore's Paradox, one should conclude that one is omniscient—that is, one should believe "if p then I believe that p, and if I believe that p, then p". I will discuss some interpretations of supposition that are *not* correct before moving towards something of a positive characterization of it.

On a "subjunctive truth" notion of supposition, $P(A|B)$ measures how strongly the agent believes that A would have been the case, if B had been *true*. This can be contrasted with what I will call a "subjunctive belief" notion of supposition, on which $P(A|B)$ measures how strongly the agent would *believe* A, if she *believed* B. If A is the event that someone wrote Macbeth, and B is the event that Shakespeare didn't write Macbeth, then for me (and I presume most people), the subjunctive truth $P(A|B)$ is quite low, while on the subjunctive belief notion $P(A|B)$ is fairly high. (I wouldn't stop believing that Macbeth had been written, even if I learned that Shakespeare hadn't written it.) The subjunctive truth notion can't be the correct account, as can be seen by cases violating the ratio account. If A is the event that I exist, and B is the event that humans have been in North America for more than 100 years then my subjunctive value for $P(A|B)$ is fairly low (if humans hadn't been in North America, presumably history would have been so radically different that I would never have come to exist), but the ratio account says that $P(A|B) = 1$, because my degree of belief in $A \wedge B$ is the same as my degree of belief in B.

van Fraassen gives an example showing that the subjunctive belief account can't be correct either, in raising worries for Brian Ellis' account of conditionals:

> Ellis' chapter on hypothetical reasoning begins with the assertion that it is the same as reasoning with conditionals. I suppose this means that if someone says 'Let us suppose that A', or 'Imagine that it had been the case that A', and goes on from there to say $B, C, \ldots, N$, then Ellis regards him as saying successively (*if A then B*), (*if A then C*), ... [van Fraassen, 1980, p. 502]
>
> We must be very careful not to call [these beliefs] something like 'the belief system this person would have if he were to come to believe that A' because then we would have difficulties with conditionals like 'If Sally were deceiving me, I would believe that she was not deceiving me (because she is so clever)' (an example due to Richmond Thomason). But of course, what we would like to learn, once we see a theory of the sort Ellis gives, is exactly what it is that a person takes to follow from a supposition. If it is not what he would believe

> if he believed the supposition, then what is it? Given his approach, Ellis will *not* answer: what would be true if the supposition were true; and given the reflections above, he *cannot* say: what would be believed if the supposition were believed. [van Fraassen, 1980, p. 503]

This is exactly the problem exploited in [Chalmers and Hájek, 2007], which arises because this account of conditional belief involves a second-order doxastic attitude. This problem was already acknowledged by Ramsey: "[The conditional probability $P(p|q)$] is not the same as the degree to which [the agent] would believe p, if he believed q for certain; for knowledge of q might for psychological reasons profoundly alter his whole system of beliefs." [Ramsey, 1926, p. 180] [Barnett, 2007] points out that some of the problems here can be traced to the distinction between "accepting p and arguing about q" and "*hypothetically* accepting p and *on that basis* arguing about q". However, it's even harder to understand the notion of having a degree of belief *on the basis of* a supposition, than to understand the notion of believing something *on the basis of* a supposition.

Thus, I don't see a good way to state just what supposition *actually* is, but I have arguments that it's not given by any of the proposals above.[2] I suspect that just as degree of belief is a primitive notion that can't be reduced to betting behavior or anything else, conditional degree of belief (or supposition) may be a primitive notion as well. In both cases, betting behavior can shed some important light on the notions, but it is not constitutive of either. For conditional probability, it seems that $P(A|B)$ ought to have some connection to the agent's disposition to accept bets on A, that will be called off if B isn't true. There is a standard Dutch book argument [Hájek, 2007a] suggesting that under this

[2]One might suspect that this fact indicates that there is *no* doxastic attitude that corresponds to the standard notion of conditional probability. This would seem surprising, because I suspect that there is *some* attitude that most readers have intuitions about, which matches the ratio account in all the examples given above. This may be the attitude of belief in an indicative conditional, but as I said earlier, I suspect that any connection between indicative conditionals and some sort of attitude must go in the other direction.

At any rate, even if there is no such relevant attitude here, I think I can let the notion of conditional probability be a purely theoretical term used in explaining many of the relations of unconditional probability to updating, confirmation, decision theory, and other notions. Rather than defining it just in terms of the ratio, as is often done, I just want to understand it to be some function that has the uses that are standardly attributed to conditional probability. Thus, if there is no relevant attitude, and conditional probability is just a theoretical notion, then section 3.3 can be taken to be a definition, rather than a series of arguments trying to relate an existing attitude to these uses.

This may make sense of the fact that students intuitively grasp that the standard ratio is a good account of conditional probability, and that this is easier for most than grasping exactly what attitude conditional probability is supposed to represent. This might mean that the mathematics that is supported by the uses to which conditional probability is put helps to clarify an attitude that was there implicitly all along, or it might mean that it helps define a natural new notion.

I think the distinction between these two possibilities is irrelevant for what I want to say.

interpretation, one ought to have $P(A \wedge B) = P(A|B)P(B)$, as well as a more complicated argument given in [Greaves and Wallace, 2006].

In chapter 7 I consider a variety of other relations that may hold (normatively) between these notions. Some have wanted to use these relations to analyze either conditional or unconditional probability in terms of the other. I suspect that the connections are not strong enough to do this—thus, I will follow [Goosens, 1979] in making the two notions equally fundamental. Just as [Eriksson and Hájek, 2007] argues that degree of belief is a primitive notion, I suggest that conditional degree of belief is as well. Thus, just as the accounts of the Ramsey test I have given fail to give an explicit analysis of the notion of conditional probability, so do the mathematical requirements on it. Thus, I suggest that rather than following Ramsey in trying to give an explicit definition of the notion of supposition, or of conditional probability, we should do something more in line with the interpretation of Ramsey given in [Lewis, 1970] and look at the roles that conditional probability plays. Thus, I will say that conditional probability is whatever mental state best plays the roles we have for it in its various applications, assuming there is one, as suggested in footnote 2. My arguments don't rule out the possibility of eventually coming up with an analysis of conditional degree of belief, but I think we can proceed better at this point by arguing about the role it plays in our theorizing than by trying to use any particular attempted analysis.

The fact that I favor this primitive notion of supposition over the subjunctive truth and subjunctive belief notions doesn't mean that those notions are useless—they just aren't the notions that correspond to conditional probability in the ordinary sense. [Joyce, 1999] suggests that the subjunctive truth notion is useful in causal decision theory, just as the ordinary notion of conditional probability is useful in evidential decision theory.[3] The subjunctive belief notion may well be important in understanding norms governing belief change.[4] But I will not focus on them here—instead, I will look at what roles conditional probability plays in our various theories.

[3]In [Meek and Glymour, 1994, p. 1009], it is argued that this subjunctive truth notion is actually just a special case of conditional probability, though one conditions on an *intervention* to produce the antecedent, rather than conditioning on the antecedent itself. If this is right, this gives one more reason to care about the notion I'm interested in here, rather than the subjunctive truth one.

[4]However, it's not totally clear that it is, because this subjunctive belief notion of supposition is still a notion relating pairs of events, and thus can't account for the cases mentioned before involving infinite sets, lack of certainty, or loss of knowledge.

3.3 The uses of conditional subjective probability

3.3.1 Updating

The most straightforward use of conditional probability is as a norm for updating beliefs—that is, if at t_1 one is certain of exactly the same things that one was certain of at t_0, together with B, then $P_1(A) = P_0(A|B)$. Of course, as I mentioned above, this particular situation is probably fairly uncommon—one rarely changes one's beliefs merely by the addition of certainty for a single event. In addition, when one gains new evidence, one's beliefs change, as does one's knowledge about one's beliefs, and similarly for even higher order attitudes. But when the events involved don't involve these attitudes, it seems natural to say that one's beliefs after learning something should aim to match one's conditional beliefs beforehand. Only in cases where the antecedent is something that is hard to suppose hypothetically does it seem plausible that these will come apart.

Another use is for so-called "Jeffrey conditionalization"—here, instead of becoming certain of B at the second time, there is some partition $\mathcal{E}$ such that the elements E_i move to specified new probabilities. The norm here is that $P_1(A) = \sum P_1(E_i)P_0(A|E_i)$. The idea for the justification here is that the evidence only directly affects the relations among the elements of $\mathcal{E}$. Thus, it shouldn't change the result of supposing any one of the elements of the partition, so $P_1(A|E_i)$ should equal $P_0(A|E_i)$ for any A. But the norm mentioned above is the unique update function that satisfies this rule, at least when $\mathcal{E}$ is finite. (I will discuss the case where $\mathcal{E}$ is infinite in section 8.1.1.)

3.3.2 Confirmation theory

Another major source of applications for conditional probability is in confirmation theory. Since the middle of the 20th century, most approaches to the theory of scientific confirmation have used probability in some way or another. The ideas of confirmation theory are intended to help us understand the notions of justification and evidence, and to explain much of the practice of science. However, as pointed out in [Maher, 1996], there is a kind of ambiguity here. We might be interested in the notion that describes the ways in which our current body of knowledge hangs together—which of our beliefs support which other ones, and which we would appeal to in defending others. Or we might be interested in a sort of prospective notion—how much would one hypothesis support a theory over another. The former seems relevant when responding to questions from a third party about why we believe the things we do. But the latter seems more relevant for much scientific practice, in deciding which experiments are more useful for confirming or disconfirming our theories. There may be further questions as well that confirmation theory might help us answer.

When we address one or another of these questions with the tools of confirmation theory, the question then arises of what sort of probability function might play a role. On the former, Glymour's "Problem of Old Evidence" becomes intractable if we appeal to the Bayesian notion of degree of belief that I mainly intend to discuss. This is because every proposition that we might appeal to as support is in some sense old evidence (since it must be known). Maher and Williamson and others give some convincing arguments that addressing this question is much easier with some sort of objective probability notion, whether it is called "inductive logic", "inductive probability", "logical probability", or "evidential probability". There may be some possibility of addressing this problem with a degree of belief function, perhaps considering the approach from Joyce and Christensen that I discuss in Chapter 5, but it would be a serious problem.

On the latter notion of confirmation, which I describe as a "prospective" notion, I think degrees of belief are much more useful. This notion describes something like how relevant an agent will think a particular piece of evidence is to her hypothesis. It measures how much her confidence in a theory is altered by a supposition of the piece of evidence, and to some extent also measures how much actually learning that piece of evidence would actually change her confidence. Since these facts are facts about a particular agent, it makes sense that the probability function to be used here is the one that compares her conditional and unconditional probabilities, rather than one based on external relations between evidence and hypotheses. For this question, the Problem of Old Evidence loses much of its force. The few apparent instances (like Glymour's example of the confirmation of Einstein's theories by the already-known perihelion shift of Mercury) should clearly be dealt with as recommended in [Garber, 1983], by pointing out that the putative evidence is not the relevant evidence at all—rather, the fact of the logical connection between Einstein's theory and the perihelion shift is the relevant evidence.

However, for both of these notions, there are generally some points of agreement. We want to say that E confirms H iff $P(H|E) > P(H)$, where P is understood in the appropriate way. (For the former account of confirmation, P is the logical or evidential probability function, perhaps conditional on some background evidence; for the latter prospective account of confirmation, P should be the agent's actual degree of belief function at the time in question.) Of course, we often want a more detailed notion of confirmation, that lets us say when one piece of evidence confirms a hypothesis more strongly than another. On this topic, [Fitelson, 1999] points out that there are many different measures of strength available. While some of these measures appeal to the unconditional probability of the hypothesis being confirmed, all of them appeal to at least some conditional probabilities, whether $P(E|H)$, $P(H|E)$, or $P(E|\neg H)$. Thus, conditional probability plays an important role in confirmation theory, whichever the correct measure (or measures) of confirmation is.

3.3.3 Decision Theory

Another use of conditional probability is in decision theory. [Joyce, 1999] argues that the way to make sense of values in rational decision making is by looking at the probabilities of various outcomes on the *supposition* that one performs a particular action. (This is to be contrasted with some previous authors that suggest there is one unified space of states with an unconditional probability distribution on it, and that the action undertaken enters in only in determining the outcome in various states, rather than shaping the probability distribution.)

Since none of the arguments I make rely on decision theory, I am not committed to one position or another here. However, Joyce does point out that the traditional distinction between "evidential decision theory" and "causal decision theory" may be able to be cashed out in whether the supposition involved is somehow "evidential" or somehow "subjunctive". If the latter is the right way to think of decision theory, then that will end up using the subjunctive truth notion of conditional probability rather than the one that I seek to describe throughout this work. But as mentioned in footnote 3, even this notion might turn out to be best understood in terms of the one that I describe.

Chapter 4

Doxastic States are Triples

Several different mathematical structures have been given the name of "probability". Probably the most well-known is that of [Kolmogorov, 1950], which involves a triple consisting of a non-empty set Ω, an algebra $\mathcal{A}$ of subsets of Ω, and a function P from $\mathcal{A}$ to $\mathbb{R}$. There have been proposals on which $\mathcal{A}$ forms a different sort of algebra, as well as proposals on which the elements of the algebra are primitives, instead of sets, so that Ω is unnecessary. Beyond these basic differences, different proposed formalisms have also imposed different constraints on the function P. There have also been proposals in which the probability function itself is not real-valued, but either more finely-grained (involving infinitesimals) or coarser (involving only an ordering, rather than real values).

My goal here in this chapter is to argue that for the purposes of understanding degree of belief, the appropriate formalism will, like Kolmogorov's, involve a triple $(\Omega, \mathcal{A}, P)$. I will leave the discussion of constraints on the numerical values of the function for Chapter 7. In my framework Ω will represent the set of epistemic possibilities, $\mathcal{A}$ will in some sense correspond to the propositions the agent can consider, and P will give the agent's degrees of belief. What exactly the elements of Ω are is beyond the scope of my discussion. It is a deep question whether epistemic possibilities are best understood as some sort of possible worlds, or as maximal consistent sets of propositions, or something else.

Similarly, there is a question as to what the appropriate bearers of probabilities are.[1] In general I will try to bracket this issue, as I think it won't affect most of the discussion. Instead, I will just neutrally call them "events", and suggest that they can be

[1]This bears a certain resemblance to the question of what the fundamental truth-bearers are, whether they are sentences, propositions, utterances, assertions, beliefs, or something else.

appropriately represented by sets of epistemic possibilities. I specifically don't mean to use the term "event" to denote some sort of concrete occurrence, or anything of the sort—it just refers to whatever the objects of the subjective probability function are, whether sentences, propositions, or something else. Although the objects of belief may well not themselves be sets of epistemic possibilities, I will argue that they correspond to such sets in such a way that two objects of belief that correspond to the same set of possibilities must get the same probabilities. Since subjective probability is a doxastic attitude, at least something about the agent's mental state should be able to single out events so that each can have a probability individually, though it's not obvious whether the agent should be able to do so in any conscious way. (See section 4.2.1 for more on this point.)

Although I don't want to assume much about what the events are, I do need to make some assumptions about their relations to one another. In particular, I assume that for any agent, the events form a boolean algebra, which I will call $\mathcal{A}$, though this algebra of events might be distinct for different agents. In particular, this means that for any two events (that is, for any two things that the agent's mental state suffices to assign a degree of belief to) there is an event corresponding to their conjunction and one for their disjunction, and that these two operations satisfy the usual rules of commutativity, associativity, and distribution. In addition there is a unary operation (called "negation") that satisfies the DeMorgan laws and double-negation elimination. And finally, there is an identity for each of the binary operations (the conjunctive identity is often called "1" or "$\top$" or "tautology" or "certainty", and the disjunctive identity is often called "0" or "$\bot$" or "contradiction" or "certain falsehood").[2] Furthermore, it is standard among most authors to presume that this algebra is not just a boolean algebra, but a σ-algebra–that is, that $\mathcal{A}$ is closed under *countable* disjunctions and conjunctions as well as finitary ones. This presumption is much harder to justify than the claim that the events form a boolean algebra, and I will raise some worries for it in Chapter 9. But most of the time I think we can also bracket this issue.

What we can't bracket is the issue of how the degree of belief function must interact with the algebraic operations on the set of events. Based on the arguments in Chapter 3, it seems reasonable to require that the degree of belief function must satisfy the basic axioms

[2]The presumption that events form a boolean algebra raises worries about the problem of "logical omniscience", in that it seems to require that the agent's mental state treat tautologically equivalent events the same way. This problem is discussed extensively in [Garber, 1983] and [Gaifman, 2004]. But I suspect that some method (perhaps a generalization of one of those two) will allow for a solution to this problem.

of probability—that is, for all x, $P(x) \geq 0$; $P(\top) = 1$; and $P(x \vee y) = P(x) + P(y)$, if $x \wedge y = \bot$. What this means for the agent is that her degrees of belief must be non-negative, that any event she is absolutely certain of must receive probability 1, and that if two events have an empty conjunction, then the probability of their disjunction must be the sum of their probabilities. I will argue in Chapter 7 for several more properties of this function, including countable additivity. (I suspect that this is the main reason most authors assume that the algebra is a σ-algebra, so that countable additivity always makes sense.)

4.1 Omitting Ω

At this point, it looks like an agent's doxastic state can be represented by the pair $(\mathcal{A}, P)$ of her algebra of events and probability function defined on that algebra. This is the approach taken in the early 20th century by Borel, and followed by some contemporary authors. (See especially [Hailperin, 1984, 1997, 2000, 2007].) However, the only argument given for this representation over a triple is that the pair somehow gives greater generality. I will argue that this generality is illusory, and that in fact there are aspects of an agent's doxastic state that the pair fails to capture, so the triple is a better representation.

Popper argues on [Popper, 1959a, p. 327] that treating the elements of the algebra as sets improperly presupposes a particular interpretation, "for in some interpretations *a* and *b* have *no members*, nor anything that might correspond to members." On the one hand, Popper's argument should only apply when the elements of $\mathcal{A}$ are themselves taken to be the objects of belief—however, as I have already stated, I am just suggesting that the objects of belief are somehow *connected* to the elements of $\mathcal{A}$, and that this connection is not the identity. After all, there are often distinct propositions that an agent regards as epistemically impossible, and these propositions will both be represented by the same element of $\mathcal{A}$, namely $\emptyset$. On the other hand, even if the elements of $\mathcal{A}$ are taken as the objects of belief, one can still construct the set Ω. Popper's axioms guarantee that the events form a boolean algebra.[3]

By Stone's Representation Theorem, [Jech, 2003, p. 81] if $\mathcal{A}$ is any boolean algebra, then we can always find *some* set Ω and a representation of $\mathcal{A}$ as subsets of Ω preserving

[3]Technically, they guarantee that the equivalence-classes of events form a boolean algebra, where two events are equivalent iff in every conditional or unconditional probability statement, switching one for the other preserves the value of the probability.

all the algebraic structure, so we can always convert any pair $(\mathcal{A}, P)$ into a triple $(\Omega, \mathcal{A}, P)$. However, one might worry that the set the theorem gives might just be a mathematical formalism. The particular set it provides is the set of ultrafilters on $\mathcal{A}$. An ultrafilter is just a maximal consistent set of events that is closed under conjunction and implication—thus, for any ultrafilter U and event E, exactly one of E and $\neg E$ is in U. If events are sentences, then this means that the states are maximal consistent sets of sentences.[4] These maximal consistent sets of sentences then seem like reasonable representations of epistemic possibilities for the agent. While they might not be the best such representation, it is reassuring that we can give a sort of meaning to them, rather than leaving them as pure mathematical constructs.

For σ-algebras this isn't quite correct, because although Stone's Representation Theorem still applies, the representation only guarantees that *finite* conjunctions and disjunctions are adequately represented by intersection and union, and not that countable ones are. In fact, the representation given by Stone's Representation Theorem *never* properly represents infinite conjunctions or disjunctions.[5] However, there are other representation theorems showing how to represent σ-algebras as algebras of sets, though the representation is more complicated. [Halmos, 1974, p. 101]

In fact, I claim that something may be lost by using the pair instead of the triple—the set Ω represents an important part of any agent's doxastic state. As I will argue in Chapter 5 and section 6.1, we need to draw a distinction between what I call "miniscule events" (ones whose probability is less than any positive real number) and ones whose falsehood the agent is certain of (which I will call "empty events"), even though miniscule and empty events will both get probability 0. One clear way to do this is by representing events as subsets of Ω, and representing events whose falsehood the agent is certain of as empty. Thus, not only is there no argument for omitting Ω from the triple, but in fact there

[4]Note that both maximality and consistency are relative to the algebra $\mathcal{A}$, rather than being matters of logic. For instance, an agent's algebra may make "The coin comes up heads" and "The coin comes up tails" inconsistent, even though as a matter of logic alone they are consistent.

A further potential advantage of representing events as sets of states is that it allows the logical relations to be non-classical. If the constraints on the probability function are connected to the operations of union and complement instead of disjunction and negation, and the states allow for an event and its negation to both be true, or both be false, then we can represent an agent who believes in a non-classical logic.

[5]I thank Eric Wofsey for explaining this fact to me—in any σ-algebra, for any countable sequence of events x_i, where each x_{i+1} entails x_i, there is an ultrafilter containing each x_i but not their countable conjunction. This ultrafilter will be in the intersection of the sets corresponding to the x_i, but not in the set corresponding to their conjunction, and thus the intersection of the sets never corresponds exactly to infinitary conjunction.

is an argument against this omission.

Considering Ω as the set of epistemic possibilities will, I think, make it more plausible that it plays an essential part in the representation of an agent's doxastic state. (For reasons of neutrality, I will refer to the elements of Ω as "states", just as I refer to the elements of $\mathcal{A}$ as "events".) These may just be maximal consistent sets of events (as given by Stone's Representation Theorem, and suggested on [Garber, 1983, p. 113]), but they may also be something more like possible worlds, or perhaps something else. Although metaphysically, the objects of the belief attitude might not be sets of epistemic possibilities, it seems that whatever they are, they must somehow be connected to these sets.

Most straightforwardly, degree of belief is a very fine-grained sort of attitude to have towards a proposition. However, there is a much coarser attitude available of deeming it possible. In order to be willing to pay for a bet that pays off only if a proposition is true, there must be some epistemic possibilities for the agent on which that proposition is true. Similarly, in order for an agent to rationally assign one event a higher probability than another (and thus be more willing to bet on the former than the latter), there must be *some* epistemic possibility for her on which the former occurs but the latter doesn't. Thus, it seems safe to suggest that events that correspond to the same set of epistemic possibilities (or states) get the same probability, and then for the purposes of probability to just replace the objects of belief by these sets of states. Many different objects of belief (whether propositions or sentences or whatever) may be represented by the same set of states.

Thus, when considering events it is often useful to just consider the set of states corresponding to the event, just as when considering states, it is often useful to just consider the set of events that are true at that state.[6] I don't want to make any assumptions about whether one or the other of these identifications is actually correct, but I claim that for the purposes of probability, it will often be fine to treat events as sets of states or vice versa, if it is convenient.

[6]Say that two states s and t are "indiscernible" ($s \approx t$) iff they are members of exactly the same events. It is easy to see that this relation is an equivalence relation. In general, I will assume that this equivalence relation is trivial. This can be arranged by using using $\Omega/\approx$ (the set of equivalence classes under the indiscernibility relation) instead of Ω. That is, if two states are elements of all the same events, they will be treated as a single state for most purposes.

4.2 Omitting $\mathcal{A}$

It is very easy to conflate what I am calling "states" and "events", or use one term to refer to both interchangeably. This is quite natural when Ω is finite, or when $\mathcal{A}$ is a σ-algebra and Ω is countable (such probability spaces are called "discrete"). In such cases, it is straightforward to see (once we have identified states according to the equivalence relation mentioned in footnote 6) that every singleton is an event. But since there are only countably many states, we see that *every* set of states is a union of countably many singletons, all of which are events. Thus, every set is an event. In addition, if we assume countable additivity, then the probability of any event is just equal to the sum of the probabilities of the singletons it contains. As a result, in discrete probability spaces, one can assign probabilities directly to the states, treating probabilities of arbitrary events (and thus arbitrary sets of states) as derivative.

However, this doesn't work in general, as some examples of uncountable abstract probability spaces will show. For instance, consider two probability spaces, in both of which $\Omega = [0, 1]$, and $\mathcal{A}$ is the σ-algebra of Borel sets (that is, the smallest σ-algebra containing all the open intervals (x, y) with $0 \leq x < y \leq 1$). It is straightforward to see that every singleton is an event in such a space, but it is also well-known that not every set of states is a member of this σ-algebra.[7] Let $P_1((x, y)) = (y-x)$ for the open intervals, with probabilities for all other sets given by the unique countably additive extension to all Borel sets. Let $P_2((x, y)) = (y - x)/2$ if $x < y < 1/2$ and let $P_2((x, y)) = 3(y - x)/2$ if $1/2 < x < y$. We can easily calculate the probability of any open interval spanning 1/2 by adding the probabilities of its part on either side.[8]

In both spaces, it is straightforward to check that the probability of any singleton is zero, even though many intervals get different probabilities on the two functions. Thus, assigning the same probability to each singleton is not a sufficient condition to guarantee that two probability functions are the same on all events. For this reason, it is important to consider probabilities as directly attaching to events, and *not* primarily to states. If the singleton consisting of a single state is itself an event, then it is common to treat this state as if it had a probability, but this is just what mathematicians call an "abuse of notation"—it

[7] This is because all members of this σ-algebra have a well-defined Lebesgue measure, and many constructions (using the Axiom of Choice) give rise to non-Lebesgue measurable sets—see below.

[8] We can also see that $P_2(\{1/2\}) = P_2(\{0\}) = P_2(\{1\}) = 0$, because these three events together form the complement of two sets whose probabilities add to 1. Thus, $\{1/2\}$ will not make any additional contribution to the probability of any interval spanning the midpoint.

is often a convenient way to simplify notation, even though it is not technically correct. In countable spaces, which are common in a lot of examples, the abuse is entirely innocent, but that doesn't mean that there is no abuse. In Chapter 9 I will consider a space in which this can't even be done indirectly, because almost no singletons are themselves events. When working with full σ-algebras this is much less common—however, if we let $\mathcal{A}$ be any complete, atomless boolean algebra, then the representation of this algebra as an algebra of sets will be a σ-algebra with no singletons.

Since events, and not states, were defined as the bearers of probability, we should remember to keep this distinction in mind. States may indirectly be assigned probabilities through their singletons, but this is only a derivative notion, and doesn't suffice to characterize the assignment of probabilities to the entire algebra.

4.2.1 Events are simple

Even when one is careful to make the distinction between states and events, it is very easy to assume that $\mathcal{A}$ is unimportant as part of the representation of the agent's doxastic state, because we may always want to deal with the complete power set of Ω. I will argue here that the events form some *sub*-algebra of the full power set of Ω, and that this subset is often proper. $\mathcal{A}$ will consist of the subsets of Ω that receive a probability.

It is a commonplace that not every set of states can have a probability. The following example was first given by Vitali to show that Lebesgue measure doesn't extend to every set of reals. We say that two real numbers in the interval [0,1] are "equivalent" if their difference is a rational number. Now, given these equivalence classes, construct a set S that contains exactly one member of each equivalence class. (Such a set exists, by the Axiom of Choice.) Vitali's argument that S is unmeasurable relied on countable additivity, and the assumption that probability is preserved under rational shifts of a set. Since the union of these countably many disjoint sets is the whole interval, and all have the same probability, this probability must be zero. However, this contradicts countable additivity, because the sum of countably many zeros is itself zero, rather than adding up to 1.

One might suspect that the work here is being done by the assumptions of uniformity under translation, and countable additivity. And it turns out [Jech, 2003, p. 130] that (assuming the consistency of certain large cardinal axioms) there are in fact probability functions that assign a value to every set when we weaken either of these assumptions.

But I claim that this fact is irrelevant for subjective probability—there really are sets of states that don't have a probability, regardless of the set-theoretic facts. Consider the set S defined by Vitali—we can't prove that such a set exists in any constructive way; it depends essentially on the Axiom of Choice. But there are uncountably many such sets that each contain exactly one member of each equivalence class (in fact, as many as there are *sets* of real numbers). But Vitali's argument shows that these sets can't all have the same probability, so if the agent is to assign probabilities to them all, then *something* must distinguish them. However, there is no property that singles out this set S uniquely, without also applying to some other set S' that intersects each equivalence class exactly once. Thus, there is a contradiction, unless something about the agent can single out each arbitrary subset of Ω.[9]

In general, the idea is that the notion of probability that I'm interested in is a psychological one, so facts about it must somehow supervene on the agent's mental states. These states might include dispositions, as well as actually realized facts (after all, even the original Ramsey reduction to betting behavior essentially involves dispositions), but this still imposes a limitation on what sets of states can constitute events. One might suspect that since an agent is finite, the events might be restricted to sets of states that can be defined by some formula in first-order logic, perhaps using finitely many particular states as parameters. This might turn out to be the right restriction, but it seems possibly too narrow. However, one thing that seems clear is that sets whose existence can only be proven using the Axiom of Choice are too complicated for any fact about the agent to single out. Thus, although I think the mathematicians of the early 20th century who opposed the Axiom of Choice were being overly psychologistic about mathematics, their restriction might be the appropriate one for dealing with a psychological notion of probability. Thus, no subset of Ω can be an element of $\mathcal{A}$ unless it can be proven to exist from ZF, rather than ZFC. (See [Jech, 2003] for more on this distinction.) There may be further relevant restrictions, but this one will suffice for most of my purposes throughout.

At any rate, just as Ω is essential for the triple to distinguish empty from non-empty events (and because I suspect no useful generality could be gained by dropping it),

[9] One difference between states and events—events must be assigned real numbers, so some fact about the agent's mind must pick out the event uniquely. However, states don't have such an assignment. The only thing the agent's mind must do to define Ω is somehow separate the epistemically possible states from impossible ones. Thus, they don't have to be individually definable, just collectively definable. The set of all epistemic possibilities must be definable, as must each set that corresponds to an event. But the individual states don't have to be.

so $\mathcal{A}$ is also essential, to distinguish the sets of states that receive probabilities from those that don't. And clearly P is essential in order to assign the probability values. Thus, an agent's doxastic state really must be represented by the triple $(\Omega, \mathcal{A}, P)$—none of the three can be dropped without missing out on some important aspect of the agent's doxastic state.

4.3 Updating the Triple

Conditionalization and Jeffrey conditionalization, discussed in section 3.3.1, are two standard ways that have been described to explain how an agent's degrees of belief ought to change. But if, as I suggest, the doxastic state is given by a triple, rather than just a function from a set of propositions to real numbers, then I should say how they affect the whole triple, rather than just saying how they change P.

In either case we could leave Ω and $\mathcal{A}$ untouched, and just adjust the probability function. However, this doesn't seem to properly represent what has gone on—learning with certainty should mean a loss of epistemic possibilities (members of Ω) Therefore, in the simple case of updating by a certainty, it seems that the natural move is to let $\Omega_1 = \{\omega \in \Omega_0 : A \text{ is true in } \omega\}$, and to let $\mathcal{A}_1 = \{A \cap \Omega_1 : A \in \mathcal{A}_0\}$.

In the Jeffrey case it is less clear how to proceed—since no certainties are gained, it's not clear how one would change Ω to represent the update. However, there is a way to induce the Jeffrey update by a kind of conditionalization on a certainty, though this event one conditionalizes on was not an element of the original algebra $\mathcal{A}$. The way to do this is to first move to a larger space $\Omega \times [0, 1]$, given by the set of pairs of elements of Ω with elements of $[0, 1]$. The new algebra is generated by $A \times [0, 1]$, for each $A \in \mathcal{A}$, together with one more event J. J will be the set of points $\{\omega, x\}$ for $\omega \in \Omega$ and $x < P_1(E_i)/P_0(E_i)$, where E_i is the unique element of the partition that contains ω. The probability function is just the product of P with Lebesgue measure on $[0, 1]$. It is straightforward to check that now, update by conditionalization on J will yield exactly the probabilities for every event that the Jeffrey update would.

This approach allows the standard conditionalization to be seen as a special case of Jeffrey conditionalization. Of course, this update process leaves some questions unanswered—for instance, what epistemic possibilities might be represented by the interval $[0, 1]$? Perhaps more problematically, is it even clear that there should be some such event J that is conditionalized on with certainty? However, the alternative would be to leave Ω

and $\mathcal{A}$ unchanged. This would draw a very deep distinction between Jeffrey updating and traditional conditionalization, which would falsify the claim that is sometimes made that traditional conditionalization is some sort of limiting case of Jeffrey update. However, it's not clear that any of these considerations are definitive about how to properly understand Jeffrey updating. Perhaps both versions of it are warranted in different circumstances.

Chapter 5

Zero and Infinity Aren't Irrelevant

Alan Hájek's "What Conditional Probability Could Not Be" [Hájek, 2003] argues that standard formulations of probability theory are flawed because they interpret conditional probability as a ratio, which is only defined when the denominator has a non-zero value. He gives examples of cases in which this condition fails, and yet we have clear intuitions for the value of some relevant conditional probability, so the analysis must be incorrect. However, some may be tempted to argue that the examples he gives are all very artificial and contrived, so that the problem he discusses isn't real. In this chapter I will defend the importance of the problem he raises, and argue that these events are important to theorizing about probability, and in particular to subjective probability. The important fact about these events is that they must have probability less than any positive real number, if they have any probability at all. At this point, I will remain neutral as to whether these are best considered as having probability 0, or by using infinitesimals of one sort or another—to avoid coming down on one side or another here I will call them "miniscule events".[1] As Hájek argues, either way, the existence of miniscule events shows that the traditional analysis of conditional probability as a real-valued ratio between two unconditional probabilities must at least be modified or supplemented, if not replaced.

However, I will first deal with some arguments that I think do *not* support conditionalizing on miniscule events. In particular, several such arguments rest on conditional-

[1] I will argue in section 6.1 that infinitesimals as they have standardly been considered are not the right way to do things, so that we will have to assign these events probability 0. I thank Laurie Paul for the suggestion that I find some term for these events at this point, rather than presuming that they get probability 0 before giving my argument, especially since I am not as strongly committed to this claim as I am to the claim that it makes sense to conditionalize on some miniscule events.

izing on contradictory propositions, or on propositions that one is certain are false, which every account of probability requires to have probability 0. I will call these "empty events", to distinguish them from miniscule events in general, and to emphasize that the set of epistemic possibilities for the agent that are consistent with these events is empty. These sorts of arguments rest on the idea that we should be able to conditionalize on any proposition whatsoever.[2] However, as part of my argument that some miniscule events should be conditionalized on, I will draw a distinction between the events that one *should* be willing to conditionalize on, and the ones that one *shouldn't*. In particular, I suggest that we shouldn't conditionalize on empty events, and we should conditionalize on all others. But all I need to establish here is that some miniscule events should be conditionalized on sometimes. One consequence of this is that the doxastic state of an agent should be represented by a triple $(\Omega, \mathcal{A}, P)$, rather than just the pair $(\mathcal{A}, P)$, so that the distinction between miniscule and empty events can be properly drawn.

But before arguing to one side that we shouldn't conditionalize on empty events, and to the other that we should occasionally conditionalize on miniscule events, I want to argue that these cases should actually influence our theory of subjective probability. One might worry that they are strange enough or far enough removed from ordinary cases that they would lead to bad theories of probability. After all, "tough cases make bad law". Our intuitions in very strange cases may not be very reliable or robust at all, which could undercut the significance of any argument based on them.

However, with conditional probability, every option that is on the table agrees numerically on every case *except* the miniscule cases. In addition to the traditional ratio analysis, and alternative accounts discussed in [Popper, 1959a] and [Rényi, 1970], there are "finitely additive conditional probabilities" [Dubins, 1975] and "regular conditional distributions" [Kolmogorov, 1950, p. 50]. If there is to be a resolution of the debate between these different accounts, it seems that arguments dealing with their numerical differences will be important.

In addition, consideration of these formalisms (which arose in response to the problem of miniscule events) has been very fruitful. They have shed light on the debates

[2] One popular formalism for conditional probability adopted by many authors who endorse conditionalization on events of probability 0 is given in [Popper, 1959a]. As his second postulate about probability, he explicitly requires that for *every* ordered pair of events A and B, there is a real-valued conditional probability $P(A|B)$. But as I will point out later, I think even Popper's approach doesn't do what some have wanted here.

about the relative priority of conditional and unconditional probability [Hájek, 2003] and finite versus countable additivity [Schervish et al., 1984], and may lead to new ones as well. So even if these are strange cases, their fruitfulness as sources of new arguments and new positions suggests that they are worthy of consideration.

5.1 Conditionalization on Empty Events

Arguments for Such Conditionalization

One argument relying on empty events is a response to what is called the "old-evidence problem" for Bayesianism [Glymour, 1980]. Traditional measures of the confirmational strength of a piece of evidence E compare the probability $P(H|E \wedge K)$ of some hypothesis H conditional on the evidence (and some background knowledge K), to the probability $P(H|K)$ conditional just on the background knowledge. However, if the evidence is already known, so that it is part of the background knowledge, then these two values are the same, and it can have no confirmational force. However, this conflicts with scientific practice. Although a relatively precise value for the perihelion shift of Mercury was well-known for several decades before Einstein came up with his theory of relativity, this value was widely seen as an important piece of evidence confirming the theory.

There are a variety of responses to the old-evidence problem, but one in particular, discussed in [Joyce, 1999, pp. 200-15] is relevant here. On this account, (at least one measure of) confirmational strength is arrived at by comparing $P(H|E \wedge K)$ to $P(H|\neg E \wedge K)$, rather than to $P(H|K)$. Since E is entailed by K, this means that we must conditionalize on the contradictory proposition $\neg E \wedge K$.[3]

Another sort of problem that might motivate conditionalization on empty events involves self-locating belief. If one supposes that a rational agent has degrees of belief represented by a subjective probability function, and that the only way she ever updates her beliefs is by conditionalization on what she has learned, then indexicals pose some interesting problems. Consider the case where one believes the self-locating facts expressed

[3] Joyce actually formulates this without the background information K, just considering $P(H)$, $P(H|E)$, and $P(H|\neg E)$. But since $\neg E$ is the negation of the evidence, it must be miniscule. One might then try to suggest that perhaps it is not empty. However, the motivation for this particular position is unclear—once one allows for cases in which one's evidence is false, it's not clear why old evidence would always have probability 1. Even in this case, this would just allow for the possibility that some old evidence could have a negation that was miniscule without being empty—there's no reason to suppose that all (or even very much) of it would have this status.

by the following sentences: "Today is Monday"; "Today is a cold day"; "Tuesday will be a hot day." At the end of the day, one goes to sleep, and then wakes up the following day, and (it seems, rationally) comes to believe: "Today is a hot day". Since one has changed one's beliefs, it seems there must have been some sort of conditionalization. The natural candidate is to conditionalize on the indexical sentence: "Today is Tuesday". This is a sentence to which one assigned a miniscule probability on Monday. However, this proposal involves conditionalization on empty events, if one does it more than once.[4]

Hájek himself also considers an argument that requires conditionalizing on an empty event. Consider a toss of a fair coin at noon. Let us suppose that the coin actually landed heads, so that at any time after noon, the (objective) chance that the coin landed tails is 0. However, Hájek suggests that even after noon, the chance that the coin landed tails, given that it landed tails, should be 1 [Hájek, 2003, p. 287], because he says that the probability of *any* event conditional on itself is 1. Presumably, he would endorse a similar claim about subjective probabilities, as well as for chances.

Reasons to Reject These Arguments

However, I believe there are very good reasons not to conditionalize on empty events. The simplest (but least telling) reason is just that no available formalism for conditioning on miniscule events has useful behavior on empty ones. Both Joyce and Hájek endorse the use either of Rényi or Popper functions for conditional probabilities. However, Rényi's theory deviates from the traditional formalism only by using a potentially unbounded measure function instead of a probability function (which is bounded by 1), and calculating all probabilities as ratios of these measures. This allows $P(A|B \wedge C)$ to be defined even when $P(B|C) = 0$, but only if the initial measure function assigns ∞ to C and some finite (but positive) value to $B \wedge C$. Empty events will still be required to get probability 0, and thus can't be conditionalized on.

Popper's account requires that whenever B is contradictory, $P(A|B) = 1$. For the Joyce/Christensen account of the old evidence problem, this means that old evidence can never be positively relevant to a hypothesis—only neutral or negative. Although this gets rid of the traditional problem of old evidence, it doesn't seem to be much of an improve-

[4]See [Titelbaum, forthcoming] for further discussion of this proposal and, I think, a much better resolution of the problem, which preserves much more of the idea that update is based on conditionalization than one might expect, without actually conditionalizing on empty events.

ment, because presumably the old evidence of Mercury's perihelion was *positively* relevant for Einstein's theory. In the situation with self-locating belief (assuming that one first conditionalizes on "Today is Monday and not Tuesday", and then on "Today is Tuesday and not Monday"), one will believe "Today is hot" on Tuesday—but one will also believe "Today is cold", as well as everything else. And in Hájek's example, although the probability that the coin landed tails, given that it landed tails, will be 1, so also will be the probability that the coin landed heads, given that it landed tails, which is a conclusion he explicitly rejects.

Thus, neither the Rényi nor Popper models support conditionalization on empty events in any interesting way. There may be other models that do, though none of the other currently available ones do, and I will argue that the search for such a model is misguided. As McGee says (despite the misleading title, "Learning the Impossible"), "There is no need for even an ideally rational agent to formulate a plan for how she would revise her beliefs upon learning that b, if she is absolutely certain that b is false." [McGee, 1994, p. 181] Of course, I am interested in conditional belief, rather than a plan for belief revision, but I think the point is the same.

As I suggested in chapter 4, we may suppose that the epistemic state of an agent at a time is represented by a triple $(\Omega, \mathcal{A}, P)$. The question then arises as to what possibilities an agent is considering when she supposes some event E. One idea would be to keep Ω and $\mathcal{A}$ fixed, and just entertain some P' chosen in accordance with some update rule, perhaps given by some notion of conditionalization or perhaps not.

However, if Ω is really supposed to represent the set of epistemic possibilities for the agent at a time, then this seems incomplete. When an agent supposes some proposition E, it seems that there should be *no* epistemic possibilities she considers that make E false. Thus, on supposing E, she should somehow use a different Ω as well, perhaps by removing all states in which E is false. This will also necessitate changes in $\mathcal{A}$ and P. In particular, it will now represent $\neg E$ by the empty set, since there will no longer be any states in Ω in which it is true. Thus, $\neg E$, as something that the agent is supposing to be false, will have the same properties as any contradiction, on this sort of model. The conditional probability $P(A|E)$ is then thought of as some sort of ratio between the "sizes" of the set of $A \wedge E$ states and the set of E states. But if E is empty, this ratio doesn't make any sense (though one could formally define it to always be 1, as in the Popper formalism.) Thus, provided that something like this picture is in play, it seems that probability conditional on an empty event can't have any interesting value, as would be needed by the supposed arguments given earlier.

Joyce already notes a problem like this when motivating the Rényi and Popper models. He contrasts conditional probability with what he calls “subjunctive supposition”. He says,

> When $P(C) = 0$ it is incorrect to read $P(X|C)$ as the probability that X *would* have if C *were* true. It is, instead the probability that X *has* if C *is* true. It is sometimes hard to get one’s mind around the idea that a person can think about what the world *is* like if a proposition that she is certain is false is in fact true.[Joyce, 1999, p. 201]

I suggest that it’s not just hard to get one’s mind around this idea, but that one shouldn’t try. There is no need to indicatively conditionalize on the falsity of something one is certain of. Joyce has just briefly conflated miniscule and empty events.

There may still be ways to deal with learning $\neg E$ if E is a certainty. After all, one can be certain of something, and then discover that one was wrong. However, the epistemic state resulting from actually learning a proposition and the probability function conditional on that proposition are conceptually distinct notions. Belief update may well be possible even without conditional probabilities. For the particular case of indexical beliefs, this is clear, because the learning going on when waking up on Tuesday is different from what goes on when one learns that one was mistaken originally about what day it was. We might proceed by appeal to our earlier beliefs subjunctively conditional on the evidence, or by some standardized repair strategy, or perhaps something non-rational.

But I would like to suggest that it makes no sense to say that we appeal to our earlier beliefs indicatively conditional on empty events. After all, what does it mean to say that something is epistemically impossible if not that I don’t make plans conditional on it? All our decisions presuppose the truth of anything we are certain of, and [Greaves and Wallace, 2006] shows how we can make sense of update by conditionalization as a sort of epistemic decision, and conditional probability as a plan for this sort of update. Since the negation of an empty event is an event that the agent is certain of, this seems to suggest that having a probability conditional on an empty event is making a plan conditional on the negation of a certainty. If this is right, then I suggest that it makes no sense to conditionalize on an empty event.

In addition, although it’s not totally clear what sort of attitude the indicative supposition involved in conditionalization is, it seems plausible that it can’t contradict one's certainties. The only way to even suppose empty propositions would be to give up certainty in their negations. This may well be possible (as we see when people manage to learn that something they were certain of is false), but it requires a change in epistemic state. If supposition is something like a restriction of the considered set of epistemic possibilities to those in the supposed event, then it doesn't make any sense to suppose an empty event.

Thus, I suggest that not every pair of events has a conditional probability—we must exclude pairs where the antecedent is empty. Popper's account does *formally* give every pair of events a conditional probability, but contradictions behave strangely, in that $P(A|C) = 1$ for any A. Thus, we can single out the events that have this property, and say that this conditional probability isn't a "real" one like the other values given by the function. His addition axiom already singles out these events as exceptions, so it doesn't seem too ad hoc to separate them out as the "bad" events. His system requires contradictions to be such "bad" events, and allows others to be as well, but guarantees that any bad event will have probability 0 conditional on any non-bad event. I suggest that the correct account of subjective probability will either not define these conditional probabilities, or at most do so in a manner like Popper's account.

5.2 Non-Empty Miniscule Events

Mathematical texts on probability make clear the distinction between *certain* events and the ones they call "*almost* certain". For instance, in a random walk on the integers with an equal probability of stepping one unit to the left or to the right at each point in time, it is *certain* that one will either reach $+1$ or -1, but it is only *almost* certain that one will eventually reach every single integer. [Feller, 1968, p. 344] I would like to suggest that although it makes no sense to conditionalize on certain falsehoods, there will be many *almost* certain falsehoods that one *should* conditionalize on. This is clearly true in many abstract mathematical examples, but I would like to suggest that it is also the case for more realistic probability spaces.

First, I will argue in more general terms that we should expect non-empty miniscule events to be plentiful, and then I will give more specific examples of such events. The general argument will serve to show that these specific examples aren't just isolated cases, and the specific examples will hopefully show that the situations they arise in are ones that we should be able to make sense of without much difficulty. Thus, I will try to stay away from examples involving throwing infinitely thin darts and similarly abstract mathematical constructs, and stick with more physically reasonable ones. Hopefully, the combination of these general and specific arguments will suggest that the mathematical nature of many of Hájek's examples is unimportant, and that more ordinary examples could be substituted instead. In later chapters, I will generally consider such mathematical examples, because they're easier to work with than more realistic cases. But the point here is just to show that such cases aren't just outliers.

General Arguments

One of the first important points in Hájek's paper is a result he calls the "Four Horn Theorem" [Hájek, 2003, p. 284] . This theorem states that if there are uncountably many mutually incompatible events, then uncountably many of them must have probability that is either zero, infinitesimal, vague, or undefined. It seems plausible that any event that isn't treated by the agent as impossible should be a relevant candidate for conditionalization, so provided that we restrict consideration to events with defined, non-vague probabilities, this leaves a lot of zeros or infinitesimals, either of which is miniscule. Thus, if there are uncountably many mutually incompatible (but not empty) events, then there will be many non-empty miniscule events.

Due to their combinatorial nature, every human language allows for infinitely many distinct sentences, even though they have a finite vocabulary.[5] This is one of the primary arguments for semantic compositionality. Although many of these sentences are in some sense equivalent ("Grass is green"; "Grass is green and grass is green"; "Grass is green and grass is green and grass is green"; . . .), there are clearly many that are non-equivalent ("My mother is tall"; "My mother's mother is tall"; "My mother's mother's mother is tall"; . . .). Thus, it seems that there are infinitely many propositions that are epistemically accessible to an agent. If we identify an epistemic possibility with its set of true propositions, then it seems likely that (at least initially) there are uncountably many such possibilities, because there are uncountably many such sets. Of course, not all of these sets will represent epistemic possibilities, because they will involve incompatible sets of propositions. But at this point, I will appeal to the *finitude* of our minds to argue for the *infinitude* of our epistemic spaces.

[5] In fact, [Langendoen and Postal, 1984] argues that human languages have not just countably many sentences, but in fact more sentences than any cardinality! However, I have no commitment to anything this strong, and actually think that many of the arguments I use implicitly force the set of sentences an individual consider to be countable.

As McGee says, "it would be foolhardy to imagine that we know so very much that we are able to rule out, with probability one, all but finitely many possible ways the world might be." [McGee, 1999, p. 257]

Defenders of regularity (see section 7.1) sometimes suggest that because of our nature as finite individuals, we can only ever consider finitely many propositions, and thus only have finitely many relevant epistemic possibilities.

> Nothing has been said about degrees of belief when the number of alternatives is infinite. About this I have nothing useful to say, except that I doubt if the mind is capable of contemplating more than a finite number of alternatives. It can consider questions to which an infinite number of answers are possible, but in order to consider the answers it must lump them into a finite number of groups. [Ramsey, 1926, p. 183]

By contrast, I suggest that the epistemic space should involve something like maximal state descriptions. I will appeal to finitude to suggest that we only recognize finitely many incompatibilities between propositions, and thus still have uncountably many sets of propositions left as epistemic possibilities.[6] This doesn't by itself mean that each of these states gives rise to a different event, (as mentioned earlier, not every set of states is an event) but it suggests a general framework in which the sheer number of states makes miniscule events common. And if this is the appropriate background epistemic space, then it seems that even in settings where only finitely many propositions are under consideration, some of them might turn out to be miniscule. This is in contrast to the general thought that miniscule events only matter when infinitely many propositions are under consideration.[7]

Of course, such arguments, being very abstract, are relatively tenuous. There may be reasons to suggest that one's epistemic space isn't defined by *maximal* truth assignments to propositions. And it may not be clear that conditionalization on these events will be interesting. So more concrete examples will be helpful as well. The main purpose of the

[6]The details of this disagreement will turn on what exactly the relations are between an agent and, respectively, states or events. As suggested in section 4.2.1, I think that agents must bear some relatively direct relation to events in order to have a probability for them, while states can be given in a much more implicit way. Thus, Ramsey's point applies may apply to events, though not to states. So I suggest that in general an agent's state space Ω may contain many more elements than her algebra $\mathcal{A}$, and these elements may be much more complex.

[7]Of course, the miniscule events will be very insignificant in making decisions—orthodox decision theory will give them absolutely no weight, unless they make infinite contributions to utility. But if two actions have the same outcome in all cases except for one miniscule case, then it may be relevant, assuming it is not empty. In such a case there may still be coherent probabilities conditional on this event, which is all that Hájek's arguments need.

preceding argument is just to suggest that these concrete examples aren't isolated accidents, but instead that we should expect these sorts of cases to arise in many circumstances.

Concrete Examples

As a first example, consider a physicist investigating a particular radioactive atom. She knows that it has a 50% chance of decaying within a time period equal to its half-life. Thus it has a 1 in 4 chance of decaying in the next time period of that same length, with similar exponential decrease thereafter. However, this probability is distributed continuously, so that the probability of decay at any specific moment in time must be less than any positive real number. One might object at this point that orthodox quantum mechanics and the Heisenberg uncertainty principle mean that it makes no sense to speak of an event occurring at a specific instant in time. However, given the conflict between general relativity and quantum mechanics, it seems that our physicist may well be more convinced that the atom will decay than that these quantum limitations are correct. If she gives any credence at all to the potential truth of some non-quantized theory in which instantaneous radioactive decays are possible, then she must allow for these miniscule events.

The previous example is in some ways reminiscent of examples with an infinitely thin dart, that current physics may make unrealistic. One may try to dismiss them thus, though we then reach a standoff between those who say these infinitary examples are too strange to base our account of probability on, and those who say that physical theory is too specific and contingent a fact to base our account of probability on. But other types of examples are possible.

As a somewhat similar example, consider a physicist who is trying to determine the value of some physical constant, like the charge on the electron, or the relative strength of gravity compared to other forces. Current theory and measurements have given fairly precise values to each of these, but they still leave an interval, and our physicist must distribute her credences across this interval. It seems that not only will many particular values have miniscule probability, but so will certain interesting *sets* of values, like her credence that this constant has some rational value, or some value expressible using only square roots of integers.

And for perhaps an even more ordinary example, consider Popper's motivation:

> A system [that can't conditionalize on events of probability zero] is not only weak; it is also for many interesting purposes *inadequate*: it cannot, for example,

> be properly applied to statements whose absolute probability is zero, although this application is very important: universal laws, for example, have, we may here assume (*cf.* appendices *vii and *viii), probability zero. [Popper, 1959a, p. 330]

Essentially, Popper argues that given the potentially infinite number of objects in the universe, if we start out with an assumption of independence in the distribution of properties among these objects, then any universal assertion that every object has some property (like, for instance, Newton's law of universal gravitation) will have miniscule probability. However, we need to conditionalize on these hypotheses in seeing how our evidence bears on them.[8]

In many ways, this final argument is reminiscent of the Joyce/Christensen argument based on the problem of old evidence. However, there are significant differences. Joyce and Christensen appeal to a seeming need to conditionalize on old evidence and its negation. However, old evidence seems best considered as a certainty, rather than just as an event of probability 1. Popper instead argues that many things we need to conditionalize on (like existential and universal statements) are "almost certainties", in a way that seems unavailable for old evidence. I think that Popper's argument is significant enough that we should consider miniscule events when theorizing about subjective probability, while the other is misguided. Empty events are not the sort of thing that we ever need to worry about conditionalizing on, and we should not be misled by the availability of formalisms for conditionalizing on *some* events of probability zero that we should thus countenance conditionalizing on *all* of them.

5.3 What we do with these conditional probabilities

[McGee, 1994, p. 197] suggests that "The problem we have been examining, how to revise one's system of beliefs upon obtaining new evidence that had prior probability

[8] [Earman, 1992, pp. 92-95] points out that Popper's independence assumption is stronger than is really plausible—it's tantamount to assuming that there can be no inductive support of one instance of a universal law by any number of other instances. A more plausible version of the independence assumption is based on the Principal Principle. Instead of assuming the instances are independent of one another in our agent's degrees of belief, we assume that they are independent of one another conditional on the value of a parameter p, giving the chance of each one. Earman shows that as long as there are a range of values for p that the agent hasn't ruled out, instances of the universal generalization will always confirm one another. This set-up no longer entails that every universal generalization has probability 0, but it does allow for them to. Additionally, if we allow a positive credence that Popper's version of the assumption is actually better than Earman's, then we still have to conditionalize (at least occasionally) on at least some miniscule events.

0, is not a problem that has any great practical significance. Whereas there are a great many epistemically possible theses that have probability 0, it seldom if ever happens that we learn one of them. To do so, we would have to observe an infinite sequence of coin tosses or note the position of a pointer with infinite precision." He says that the purpose of understanding this sort of conditionalization is just to improve our understanding of the logic of conditionals, following Ernest Adams. It smoothes out discontinuities that arise as values tend towards zero.

Since there are some worries about Adams' account of conditionals, and probability logic is not a universally accepted program, there might be worries about this justification for caring about conditioning on miniscule events. But I think some of McGee's earlier caveat may rest on a conflation of subjective probability and objective chance—to say that "it seldom if ever happens that we learn one of them" seems to presuppose that we will never be misguided in our assignments of miniscule probability. Even if we can't observe infinite sequences of coin tosses, or positions of pointers with infinite precision, there may be other events to which we have assigned probability 0 that we can observe more easily. We might rationally believe that such an event has objective chance 0 and thus assign it credence 0, but if the chance turns out to be positive (or even quite large), then it should not be unexpected that we learn something with probability 0.

However, I suggest that there is another, more important motivation, which came up for Karl Popper. In addition to giving the probability of conditionals (perhaps) and telling us how to update our beliefs, conditional probabilities play a role in confirmation theory. Although we may not ever observe directly a miniscule event, we often have to consider experiments that have many miniscule outcomes, and we'd like to understand how the results of these experiments bear on our hypotheses, which themselves might be miniscule. Though we might not be able to observe the outcomes exactly, or learn the hypotheses with certainties, various accounts of confirmation theory will still appeal to $P(E|H)$ or $P(H|E)$ in determining how much these experiments can confirm these hypotheses.

Thus, we should see how miniscule events affect conditional probability when we consider the mathematical constraints that hold between conditional and unconditional probability.

Chapter 6

Alternative Accounts of Miniscule Events

Many previous authors have discussed the problem of how to conditionalize on miniscule events. I will consider here several proposals in order to reject them, before moving in the next few chapters towards a positive account of how best to deal with this problem.

6.1 Infinitesimals

One popular approach to miniscule events has been to say that they don't actually get probability 0, but rather get infinitesimal values, as described by Abraham Robinson and others. These infinitesimals are part of a theory known as "non-standard analysis", and the numbers including the standard reals and the infinitesimals are sometimes called "hyperreals".

Non-standard analysis is an application of the compactness theorem of first-order logic. This theorem says that a first-order theory is consistent iff every finite subset is. We can take the standard first-order theory of the real numbers and add a single constant c and the sentences $c > 0, c < 1, c < 1/2, c < 1/3, \ldots$. Every finite subset of this theory is consistent, with a model consisting of the real numbers, with c interpreted as some sufficiently small positive number. Thus, the entire theory is consistent, and the model must satisfy all the standard first-order properties of the real numbers, but must also interpret c as some number greater than 0 but less than any positive real number. c is thus said to be an infinitesimal.

Because this model satisfies the standard first-order theory of the real numbers, there are also other infinitesimals $2c$, $c/100$, $\sqrt{c}$, and c^n, among others, and these satisfy the expected properties for arithmetic and ordering. (For more details see sources like [Goldblatt, 1989] and [Robinson, 1996].)

The suggestion is that the probability function can take values in some such non-standard model containing infinitesimals, and thus miniscule (but non-empty) events can have infinitesimal values. Since the standard first-order theory of the reals applies here, we can divide by these non-zero values without any special trouble. One famous endorsement of this view is in [Lewis, 1980, p. 268]: "Many propositions must have infinitesimal C-values, and $C(A/B)$ often will be defined as a quotient of infinitesimals, each infinitely close but not equal to zero." However, there are serious problems with this proposal.

On this approach, instead of assigning values to every event from the real interval $[0, 1]$, a probability function assigns values from the *hyperreal* interval $[0, 1]$. Using a variety of arguments for regularity (which I examine and reject in section 7.1), it is then required that only empty events get probability 0, and that all other events get some non-zero hyperreal, which may be infinitesimal or not. Conditional probabilities are then just calculated using the ratio formula, which will always give a well-defined hyperreal in the interval $[0, 1]$, as expected.

What I think is the most decisive argument against the use of infinitesimals comes from [Williamson, 2007]. In this paper, Timothy Williamson considers two separate unbiased coins, both flipped infinitely many times, once a second, with the second coin starting one second before the first. On the one hand, it appears that the probability that the first coin comes up heads every time (call it x) is equal to the probability that the second coin comes up heads every time, because both coins are fair. This should also be the probability of the second coin coming up tails on the first flip, but heads thereafter, again because the coin is fair. Thus, the probability that the second coin comes up heads on all flips after the first is $2x$.

But since this requires getting heads at all the same times that the first coin is flipped (because the second coin starts one second earlier), it seems that $2x$ should equal x. But because the hyperreals satisfy the same first-order properties as the reals, the only solution to this equation is $x = 0$. Williamson also gives a second argument based only on a comparative rather than quantitative notion of probability, to show that the probability of an infinite sequence of heads is no more probable than the empty event. Thus, even a non-Robinson type of infinitesimal will run into problems with this case, assuming that some ordinary facts about ordering of probabilities hold. This doesn't prevent their helping with some other miniscule events, but since this is already an example of a non-empty but miniscule event, I think this shows that infinitesimals will not do what is needed to establish regularity in general for any sort of subjective probability function.

One might worry that Williamson makes some illicit assumptions about which sequences of heads should be equiprobable. However, I think that further consideration of how infinitesimals work will show that they just aren't up to the task in question. They may help out for certain probabilities in certain spaces, but as Williamson suggests, there will still have to be some events we regard as possible and yet assign probability 0 to. The reason is because of the construction of the hyperreals. If we are careful with the construction, we can show that every hyperreal is close to a "hyperinteger"—but this will include $1/c$ for any infinitesimal c. The relevant hyperinteger will have to be larger than any standard integer.[1]

Therefore, there is something quite unnatural about trying to use the hyperreals to deal with probabilities of sequences that are properly indexed by the natural numbers, rather than the hypernaturals. Since this is exactly the situation that Williamson is investigating, it makes sense that hyperreals won't do the job here. Plenty of other circumstances also would result in the same sort of mismatch between the mathematics used for the probability function (where there are hyperreals and hypernaturals) and the mathematics used to describe the space (where we only have standard naturals, or only standard reals). If (as in Chapter 8) we are interested in a point with a uniform distribution over the surface of a standard sphere, we might wonder what probability we should assign to the point lying exactly on the equator. If the sphere has real (rather than hyperreal) coordinates, then there's a mismatch if we try to use hyperreals to give the probability. On the other hand, if we use hyperreal coordinates, then we might either be interested in the probability that the point lies in a certain infinitesimal strip around the equator, or perhaps some wedge, or some other shape with infinitesimal thickness that contains the equator. Any of those will get a non-zero hyperreal probability, but the equator itself will have to have a probability less than any of those hyperreals, and thus have probability 0. Thus, we either get some

[1]There is also other construction taking the ultraproduct of a set of models, using a non-principal ultrafilter on the natural numbers. In either case, we replace the reals with the hyperreals, but we also replace the naturals with the hypernaturals.

mismatch (which suggests that there will be a further problem of the sort Williamson raises) or there will still be non-empty events that must get probability 0, despite the presence of infinitesimals.

A simpler worry one might have about this proposal is that it seems to cause trouble for any account of probability in terms of betting ratios and the like. After all, bets are necessarily phrased in terms of real amounts of money, and in fact it might even be a conceptual requirement that the quantities be real rational numbers! For any two agents with distinct real probability assignments for a particular event, the betting interpretation gives some particular rational bet that one agent will accept and the other reject. Nothing like this seems to be available for two agents whose credences are given by two hyperreal numbers that differ only by an infinitesimal.

Of course, as I mentioned in Chapter 3, there are severe worries about the betting interpretation of probability. But this still suggests a related concern no matter what kind of mental state probability is—what precisely distinguishes the mental states of two agents that assign probabilities that differ only by an infinitesimal? Are there any dispositions that one will have and the other lack? Especially since the definition of the hyperreals is non-constructive, it's unclear exactly what facts about an agent could pick out any one particular infinitesimal rather than another.

One might suggest that betting behavior itself only defines probabilities up to the nearest real number—what defines the infinitesimal part is how conditional probabilities behave. In some sense however, this seems to reverse what is supposed to go on. The infinitesimals were supposed to define the conditional probabilities, while this suggestion lets the conditional probabilities define the infinitesimals. There is in fact a way to construct infinitesimals from the overall conditional probability function [McGee, 1994], but I think that such an approach is better considered as an approach using Popper functions (discussed in the next section) than an approach using infinitesimals.

Additionally, if my arguments from Chapter 8 are correct, then at least some pairs of events in some spaces should have conditional probabilities that depend on what set of alternatives are being considered for the antecedent. But assigning particular hyperreals to each event will prevent this possibility. Thus, if my arguments are correct, infinitesimals will be insufficient to explain conditional probabilities.

Of course, all this still allows that some miniscule events might be assigned infinitesimals in some way. It just means that for at least some pairs of events, their conditional probability isn't determined by the relevant ratio of infinitesimals, or else the infinitesimals are of a very different sort than the Robinson ones that are standardly used. Perhaps there are arguments for such a position, but they will look very different from the arguments in favor of the normal use of infinitesimals, and it seems fruitless to consider such a possibility given that there are no other relevant mathematical notions of infinitesimals available for use.

6.2 Popper

Another popular alternative for giving probabilities conditional on miniscule events is that given by Karl Popper in [Popper, 1959a] and elsewhere. This alternative has also been explicitly supported by [Hájek, 2003] and many others. On this picture, although the ratio analysis runs into problems with zero denominators, this is no problem for conditional probability, because conditional probability is taken as the fundamental notion. Kolmogorov's axioms are replaced by a different set of axioms, taking conditional probability as the primitive notion, and unconditional probability is just defined as probably conditional on a tautology.

These axioms have some additional mathematical elegance over Kolmogorov's axioms. They just require an uninterpreted algebra of events with a binary and unary operation, and they allow the definition of an equivalence relation between events that can be replaced for one another in any probability statement while preserving the values of all conditional probabilities. It can be proved that the quotient of the algebra by this equivalence relation gives a boolean algebra, with the standard interpretation of the unary operation as negation, and the binary operation as conjunction.

However, this added elegance is primarily mathematical—as I argued in Chapter 4, there are reasons we should already have for taking the algebra of events to be an algebra over some underlying set, and Popper's axioms don't manage to ensure that *all* logical relations are properly respected by the probabilities. There are advances on this latter point made in [Field, 1977] and [Roeper and LeBlanc, 1999], but it is at any rate not clear what advantage there is to using a probability theory that has these extra mathematical features.

The major advantages of this theory are in avoiding the problem of conditionalizing on miniscule events, and (according to some authors) making conditional probability the fundamental notion. However, this latter point may also be seen as a liability—although logical probabilities must take conditional probability as the fundamental notion (see section 2.2), this doesn't seem to be true for degrees of belief. As pointed out in [Maher, 2006b, p. 514] in drawing this distinction, "Degrees of belief are relative to a person and a time but not to evidence." Using Popper's account would reject this, by saying that all degrees of belief are really conditional in this sense. Taking such a position would thus mean that we're somewhat mistaken in our standard understanding of degrees of belief. But there may be arguments in favor of this position—this is what [Hájek, 2003] claims to give. However, the arguments there proceed primarily by showing problems with alternate conceptions of conditional probability. Since the account I give in later chapters avoids the problems discussed there, this undermines Hájek's arguments in favor of Popper's account.

6.3 Lexicographic Probabilities

Another approach that is sometimes taken is to consider a *sequence* of probability functions rather than just a single one. The first one in the sequence is taken to represent the agent's actual degrees of belief, while the later ones somehow represent "backups". To update this sequence of functions on learning some proposition B, the agent simply removes all functions on which $P(B) = 0$, and replaces all the others by their version conditionalized on B, as in the update for a standard probability function. A natural candidate for the conditional probability $P(A|B)$ is then just the standard $P(A|B)$ in the first function on which $P(B) \neq 0$. Thus, on this account, an event has miniscule probability iff its probability in the first function is 0, but since it might not be 0 on all of the functions in the sequence, we can represent the fact of its subjective possibility, and also define probability conditional on it.

6.4 Relations among the alternatives

There are a series of interesting relationships among these various alternate accounts of conditionalization on miniscule events. Perhaps the most complete source for these relations is [Halpern, 2004]. In this manuscript, Halpern shows that in some sense, Popper functions can all be represented by some sort of lexicographic probability function, and a more general class of lexicographical probability functions can all be represented as hyperreal-valued probability functions. In addition, when we look at certain natural sorts of equivalence that might hold between distinct probability functions of the same class (for instance, all the infinitesimals might be scaled up by some infinitesimal amount, or all probability functions in a lexicographic sequence might be duplicated) we find that the representations described are isomorphisms. That is, non-equivalent functions of one class get mapped to non-equivalent functions of the other, and there is a way to find a function of either class corresponding to any function of the other class.

Many of the problems argued above for infinitesimals are in some sense problems of being too fine-grained. That is, nothing could fix which particular infinitesimal was the relevant one, it seemed that probability functions that differed only by an infinitesimal value would have no different consequences for decision or confirmation, and hyperreals were too rich a field to use for dealing with objects coordinatized by standard reals. Thus, we might suspect that one of the other alternatives described here does a better job. The fact that these independently-motivated formalisms have equal expressive power also suggests that there might be something right about using one of these notions.

However, I will give arguments in the next two chapters that the correct account of conditional probability will have a property that none of these accounts do, namely that $P(A|B)$ depends additionally on a third parameter. There is a way to represent this that bears some similarity to the lexicographic probabilities, but it is importantly distinct. Where the lexicographic probabilities use a set of probability functions, and appeal to the *first* one where $P(B)$ is non-zero to give the value for $P(A|B)$, I will instead use a set of probability functions, and appeal to the *relevant* one to give the value for $P(A|B)$. Instead of indexing these probability functions in some sequential order, I will index them by partitions $\mathcal{E}$ of the probability space.[2] Each of these functions will be something like a Popper function, giving both conditional and unconditional probability values, but the only conditional probability values included will be probabilities conditional on elements of $\mathcal{E}$. These probability functions will all share the same unconditional probabilities for every event, and differ only in which conditional probabilities are defined, and what their values are. The constraints relating these conditional and unconditional probability values will be discussed in the next chapter.

[2] A partition of a probability space $(\Omega, \mathcal{A}, P)$ is just a collection $\mathcal{E}$ of elements of $\mathcal{A}$ such that every element of Ω is in exactly one element of $\mathcal{E}$.

Chapter 7

Further Conditions on Subjective Probability Functions

As discussed in Chapter 3, there are a variety of arguments for the basic tenets of probabilism—that is, the claim that beliefs come in degrees, and that these degrees satisfy the basic Kolmogorov axioms for probability—that is, that the probability of a tautology is 1, that all probabilities are non-negative, that $P(A \vee B) = P(A) + P(B)$ for mutually exclusive A and B, and that $P(A|B)P(B) = P(A \wedge B)$. I have also argued in Chapter 4 that the best way to represent an agent's doxastic state is with a space Ω of possibilities, and an algebra $\mathcal{A}$ of subsets of this space, with the conditional and unconditional probability functions being defined on the algebra. In this chapter I will consider several further apparent rational requirements on these functions, arguing against some and for others. Several of these further requirements will then form the basis for the argument I give in the next chapter that the conditional probability function must take as a third parameter a partition of Ω, and not just two elements of $\mathcal{A}$.

7.1 Regularity

Some have argued that a rational Bayesian agent should always assign a non-zero probability to any event other than a logical contradiction. A function that only assigns 0 to the empty set is standardly called *regular*.

An apparent objection to regularity as a norm for rational agents appeals to the possibility of omniscient agents. Such agents, it seems, really should assign probability 0 to every false claim, and not just to logical falsehoods. However, this argument ignores the fact that probabilities are assigned to elements of $\mathcal{A}$, which are subsets of Ω, rather than some other notion of proposition. For an omniscient agent, it would seem that there really is only one epistemic possibility, so that Ω is a singleton. Thus, the only elements of $\mathcal{A}$ would be the empty set, and this singleton! Thus although such an agent should still draw distinctions between propositions by using possible worlds, or perhaps something else more fine-grained than her epistemic possibilities, each such proposition will correspond to either the empty set or the singleton, and thus get probability 0 or 1. Thus, the possibility of omniscience is no objection to the claim that every non-empty event should get non-zero probability.

However, I think there are other arguments against regularity that work better, and I think that none of the arguments in its favor work. In Chapter 5 I give the argument that there are in fact events (which I call "miniscule") that ordinary rational agents should assign probabilities smaller than any positive real number. In Section 6.1 I give the argument that these events in fact get probability 0, rather than an infinitesimal value, or something else. But here I will focus on rebutting the positive arguments in favor of regularity. If the arguments in favor of regularity fail, and there are problems with each of the theories that has supported regularity, then it's clear that we should not accept it. (However, see Chapter 9 for an alternate approach, which is more radical, but may be able to restore something like regularity.)

The most basic version of the argument is some variation on the claim that probability 0 can't mean anything except for epistemic impossibility. However, the formalism of using a triple $(\Omega, \mathcal{A}, P)$ to represent an agent's doxastic state already gives a representation of epistemic impossibility, when a proposition is represented by the empty event. Since this formalism allows for non-empty events to also get probability 0, this first impression in favor of regularity is defeated. This argument can be sharpened by giving a sort of Dutch book-type argument—if an agent assigned probability 0 to a non-empty event, then she would be willing (or at least think it fair) to give away for free a bet that she thinks she might have to pay out. However, this depends on the assumption that agents actually judge it fair to buy and sell bets exactly at the price given by their probability, rather than merely judging it favorable to buy them at lower prices and sell them at higher prices.

Another argument suggests that a norm of update by conditionalization would prevent an agent with a non-regular probability function from ever changing her mind about these assignments of zeros. Given that rational agents that are non-omniscient seem perfectly capable of recognizing their own fallibility, it seems that they should be required to acknowledge this, and therefore never assign a probability value that they can't later undo.

> I should like to assume that it makes sense to conditionalize on any but the empty proposition. Therefore I require that C is *regular*: $C(B)$ is zero, and $C(A/B)$ is undefined, only if B is the empty proposition, true at no worlds. ... The assumption that C is regular will prove convenient, but it is not justified only as a convenience. Also it is required as a condition of reasonableness: one who started out with an irregular credence function (and who then learned from experience by conditionalizing) would stubbornly refuse to believe some propositions no matter what the evidence in their favor. [Lewis, 1980, p. 267]

However, it's not clear why an agent always has to update by conditionalization. After all, update by conditionalization only exacerbates the problem by setting the negations of all learned propositions to probability 0. If the agent was ever mistaken in one of these updates, then undoing the damage can't be done by conditionalization, so she must at least sometimes be permitted to use some other procedure.

Now, this isn't such a worry for Lewis—in the quoted passage he is only arguing for regularity for "initial credence functions", rather than the ones held by agents that have actually learned some information. He is not concerned with the problem of an agent having mistakenly "learned" some proposition, but rather by an agent having, *in the absence of any information whatsoever*, set her credences in such a way that she can't change some feature of them in light of the facts. If the agent really had done this, there would be a problem for the agent, but Lewis is wrong in his claim that irregular credence functions are stubborn in the sense described.

It's true that updating by conditionalizing on an event of non-zero probability can never raise an event of probability 0 to a non-zero value. However, it's not clear why Lewis thinks that agents would only learn events they initially assigned non-zero credence to. If the agent really assigned probability 0 to an event A she considered subjectively possible, then it should be possible for A to be true, and for her to learn A. If the agent has a well-defined value for $P(A|A)$ (which presumably should be 1), then this particular learning event would falsify Lewis' claim that irregular credence functions are stubborn. Additionally, there are plenty of other propositions that would suffice—she could learn something that entailed A, or she might learn some other event of probability 0 that was consistent with A, but didn't entail it. It's true that if $P(A) = 0$ and $P(B) \neq 0$, then $P(A|B) = 0$, but if $P(B) = 0$, then the standard formula $P(A \wedge B) = P(A|B)P(B)$ gives no constraint to $P(A|B)$. In the next chapter I will give my positive account of these conditional probabilities and show that very often, $P(A|B)$ will be non-zero (and can in fact take on any value between 0 and 1), even though the specific values can't be calculated in many cases. Thus, Lewis' argument is premised on a mistake, that an agent who assigns something probability 0 can never conditionalize on an event of probability 0. If this argument is part of an argument against assigning probabilities conditional on events of probability 0, then it begs the question, and if it is part of an argument against assigning probability 0 to non-empty events, then it has a false premise.

Regularity is also appealed to occasionally as part of a response to Glymour's "Problem of Old Evidence." The response suggests that the old evidence problem should never arise, because a rational agent would never actually assign anything probability 1—she would only update by Jeffrey conditionalization, and thus there would always be room for confirmation. However, I think this response is inadequate, and if there are independent arguments against regularity, then it can't even properly work. (My primary discussion of this problem is in section 7.3, and in particular in footnote 6.)

So far, I have been considering the suggestion that an agent should never assign probability 0 to any event that she thinks is subjectively possible. However, it is also possible to think of regularity as a requirement that an agent should not deem subjectively possible any event that she assigns probability 0 to. I consider this version of regularity in Chapter 9.

A weaker principle of this form is discussed extensively in [Shafer and Vovk, 2005], under the name "Cournot's principle". According to this principle, "An event with very small probability is *morally impossible*; it will not happen." (p. 9) They cite Cournot as saying "The physically impossible event is therefore the one that has infinitely small probability, and only this remark gives substance—objective and phenomenal value—to the theory of mathematical probability." The idea is that without this principle, the theory of probability makes no observational predictions, and therefore (from something like a positivist point of view) has no content.

Something like this principle is necessary for any application of the law of large numbers—Bernoulli's theorem proves that in a large enough sequence of trials, the probability of the relative frequency differing by more than a small amount from the probability is extremely small. By making it small enough, we can apply Cournot's principle, and therefore infer that the probability *is* in fact close to the observed relative frequency. Similarly, other statistical tests can be directly used to infer probabilistic claims about hypotheses.

However, this principle is mainly necessary only to give meaning to an objective notion of probability. There is no need for such a principle in subjective probability—on the one hand, we already have notions of betting and certainty that give the probability its meaning, and on the other, we don't expect the agent to be infallible. Even a rational agent might be misled into judging a true proposition to be miniscule, especially if there are infinitely many alternatives..

But even in the objective case, such a principle seems controversial. After all, one fact about small probability events is that they do in fact occur! d'Alembert seems to have claimed [Shafer and Vovk, 2005, p. 9] that a run of a hundred heads in a sequence of coin flips will never occur, appealing to something like Cournot's principle. However, whatever sequence does occur when flipping a coin will have the same probability, so Cournot's principle should apply equally to it. Thus, it seems that even for objective probability, it might be better to appeal to a different principle to give content to the notion, perhaps following [Deutsch, 1999] or something similar.

7.2 Countable Additivity

It is one of the standard requirements of a probability function that it satisfy finite additivity—that is, that if events A and B are known to be mutually exclusive, then $P(A \vee B) = P(A) + P(B)$. A further constraint that is often proposed is that this should hold not just for pairs of events (and other finite collections, which follows from the fact for pairs) but for countably infinite sets of events as well. That is, if A_i are events known to be mutually incompatible, then $P(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} P(A_i)$.[1] This principle was endorsed by Kolmogorov primarily for its mathematical convenience, but it was opposed by de Finetti.

The primary argument against countable additivity has always been that it rules out certain probability distributions that intuitively seem like they should be possible. In particular, de Finetti argued that it should be possible for a rational agent to be uncertain about the value of some number n, being certain that n is a natural number, but not believing any value any more than any other. In this situation, finite additivity shows that

[1] If $\mathcal{A}$ is not a σ-algebra, then this must be restricted to the case where $\bigcup_{i\in\mathbb{N}} A_i$ is itself a member of $\mathcal{A}$.

the agent's credence in any particular value cannot be positive—if it were greater than $1/k$, then the probability that the number is between 1 and k would be greater than 1, which is impossible. Thus, in this situation the agent's credence in each value should be zero. However, this results in a violation of countable additivity, since the credence that the value is some natural number is 1 (because the agent is certain of this fact), which is greater than the sum of all the credences of the individual values (which is 0). Not only does this distribution perhaps seem intuitively permissible for a rational agent, but it is also in a natural sense the limit as k goes to infinity of the distributions that assign all credence uniformly to the finite set of natural numbers up to k. The defender of countable additivity seems to suggest that each of these finite uniform distributions is permissible for a rational agent, even though their limit (which is finitely, but not countably additive) is not.[2]

Nevertheless, many people have argued in favor of countable additivity as a further requirement for subjective degree of belief functions. [Williamson, 1999] points out that many of these arguments (like Kolmogorov's) depend just on the mathematical convenience of this further postulate, and the fact that it limits us to distributions that are well-behaved in various ways. But he argues further that there is a foundational justification for countable additivity, in the form of a Dutch book. This Dutch book basically takes the same form as the Dutch book for finite additivity, with the agent selling a bet on each disjunct and buying a bet on the disjunction, for a guaranteed loss if the agent values the disjunction higher than the sum of the disjuncts. However, as discussed in Chapter 3, Dutch book arguments have various problems, and this one has even more. First of all, nothing about this argument seems to require that the set of events being added together is countable—thus, the argument should suggest that probability distributions must be "perfectly additive", meaning that for any disjoint set of events, all but at most countably many of them have probability 0, and the probability of the disjunction of all the events must just be the sum of the probabilities of the countably many non-zero ones. However, just as countable additivity rules out a uniform distribution on a countable set (like the natural numbers), perfect additivity rules out a uniform distribution on any infinite set, including the unit interval of real numbers between 0 and 1. But this distribution seems much more natural than the uniform distribution on the natural numbers (for instance, it

[2] A further approach to this situation is described in [Bartha, 2004]. However, this approach is based on an approach to subjective probability that takes *ratios* of probabilities to be fundamental, rather than conditional or unconditional probabilities. This approach also allows for particular probabilities and ratios of probabilities to be undefined.

is easy to calculate the probability of any definable subset of this interval), so this Dutch book argument seems to prove too much.

A natural diagnosis of what goes wrong with this argument is that the Dutch book involves infinitely many bets. Thus, the argument only works if we assume that a rational agent that judges each of an infinite set of bets to be fair must also judge the entire set to be fair. But if this is the case, then we can get absurd results. [McGee, 1999] describes an "airtight Dutch book" that guarantees a loss for *any* agent. There is an infinite sequence of coin flips, and each bet is a bet that the following flip will come up heads, conditional on all previous flips coming up heads (but not all infinitely many flips coming up heads), each at slightly more than double the stakes of the previous ones. The agent is guaranteed to win some initial sequence of these bets, lose the next (whose stakes are slightly higher than the sum of all the earlier bets), and the rest will be called off. McGee also shows that a version of this Dutch book can always be set up unless the agent has a bounded utility scale, or has only finitely many events with non-zero credence. The appropriate response here seems to be to deny that being susceptible to a Dutch book involving infinitely many bets is irrational. But then, we lose Williamson's argument for countable additivity.

[Howson, 2008] continues de Finetti's arguments to the effect that we should reject countable additivity. In particular, he suggests that if the arguments are based on Dutch books, then they are not to be accepted. Instead, the argument for probabilism that he accepts is the Cox-style argument given in Chapter 3, and he points out that there is no natural generalization of this argument that gives countable additivity. I agree with Howson's points that Dutch book arguments are not to be trusted, and with his interpretation of de Finetti as arguing that mathematical convenience should not be the shaper of the theory of rational degree of belief—this is supposed to be a substantive theory, and it is only by investigating its subject matter (degree of belief) that we can find the appropriate constraints.

But all this doesn't yet rule out countable additivity as a constraint. The only argument against it is that it rules out certain distributions that somehow seem natural, although there are no strong arguments that this distribution really should be permissible.[3] The arguments in its favor that I have given so far are either inconclusive or depend on

[3]One might similarly question finite additivity, because it rules out a distribution on which $P(A) = P(\neg A) = 1/3$ as suggested in [Weatherson, 2003] for situations in which an agent happens to lack much evidence either for or against A. This argument is supposed to motivate an intuitionistic version of probability theory, but it could also be developed in a way that just questions additivity.

questionable premises about Dutch books. However, I will eventually argue that countable additivity is an appropriate constraint for rational belief, because it follows from the principle known as conglomerability that I will discuss next.

7.3 Conglomerability

The strongest version of conglomerability is the following:

For any partition $\mathcal{E}$ of some event B, and any event A, $P(A|B) \leq \sup_{E\in\mathcal{E}} P(A|E)$.

Substituting $\neg A$ for A, we see that it also entails $P(A|B) \geq \inf_{E\in\mathcal{E}} P(A|E)$. An equivalent formulation of conglomerability is given in Appendix A. That formulation requires that the unconditional probability $P(A \wedge B)$ is equal to an integral over B of certain conditional probabilities of A. It is a generalization of the standard "law of total probability", which asserts that if $\mathcal{E}$ is a partition of B into *finitely* many sets, then $P(A \wedge B) = \sum_{E\in\mathcal{E}}(P(A|E)P(E))$. For finite partitions, this result follows from finite additivity, and the fact that $P(A|E)P(E) = P(A \wedge E)$. If $P(B)$ is non-zero, then dividing both sides by $P(B)$ gives $P(A\wedge B)/P(B) = \sum_{E\in\mathcal{E}} P(A|E)(P(E)/P(B))$. But since $P(A\wedge B)/P(B) = P(A|B)$, and since $P(E) = P(E\wedge B)$, so $P(E)/P(B) = P(E|B)$, this expresses $P(A|B)$ as a weighted average of the $P(A|E)$, where the weights are $P(E|B)$—thus, it guarantees that conglomerability holds for this partition, because a weighted average of some numbers can never be strictly greater than all of them. Thus, conglomerability is only an additional constraint when we consider infinite partitions $\mathcal{E}$.

There are also weaker versions of conglomerability one could add as constraints, in case the full version is too strong. One way to weaken it is to require it only in the case where $P(A|E)$ is constant for all $E \in \mathcal{E}$. In this case, it says that if $P(A|E)$ is constant over some partition $\mathcal{E}$ of B, then $P(A|B) = P(A|E)$ for any element E. Another way to weaken conglomerability is to require it only when B is itself the whole space, so that $P(A|B) = P(A)$. This version states that $P(A) \leq \sup_{E\in\mathcal{E}} P(A|E)$.[4] Combining both of these weakenings, we get the statement that if $\mathcal{E}$ is a partition of the whole probability space (so that the agent is certain that exactly one of the events $E \in \mathcal{E}$ will occur), and if $P(A|E)$ is constant for all $E \in \mathcal{E}$, then $P(A)$ must equal this same constant value.

[4] An argument that this version is in fact weaker than full conglomerability appears on [Hill and Lane, 1986, p. 56] However, [Kadane et al., 1986, pp. 67-8] shows that the weaker version implies the stronger version for all B such that $P(B) > 0$.

This version just identifies two distinct notions of probabilistic independence. We can say that A is independent of the elements of a partition if $P(A|E)$ is constant over the elements of the partition—this is independence in the sense of it not mattering which specific event is conditionalized on. But the more standard notion of independence for a particular event says that A is independent of E iff $P(A) = P(A|E)$. The weak form of conglomerability just states that if A is independent of the partition in the first sense, then it is independent of each element of the partition in the second sense. This weakest version is the only one that I will require in the argument of the following chapter, but I think there are good arguments in favor even of the strongest version.

The principle of conglomerability is a synchronic principle of rationality that is closely related to a diachronic principle called "reflection" by van Fraassen. [van Fraassen, 1984, van Fraassen, 1995] This principle says that if the agent knows at time t_0 that at some point t_1 in the future, when she is operating rationally and is in at least as good an epistemic state, she will assign $P_1(A) = x$, then she should already assign $P_0(A) = x$. This seems plausible, because the agent knows that her future self can only be in a better epistemic position with regards to A than she is right now, and so she should defer to this self in assignment of probabilities.

Consider the special case where the only information learned between t_0 and t_1 is which element of $\mathcal{E}$ is actual, and assume that the conditional probabilities $P_0(A|E_i)$ are all equal. If we assume further that the agent in this situation knows that all this is true, and also knows that she will update her degrees of belief by conditionalization (so that $P_1(A) = P_0(A|E_i)$, where E_i is the actual element of $\mathcal{E}$), then reflection requires exactly what conglomerability requires, that the unconditional $P_0(A)$ equal $P_1(A)$, which is $P_0(A|E_i)$.

Because of the requirement of these extra factors, and the difficulties involved in specifying exactly what it means to be in "at least as good an epistemic state" (especially when it comes to cases of self-locating information, as in [Elga, 2000], and various responses, like [Titelbaum, forthcoming]), reflection generally says less than conglomerability. But their similarities suggest that similar defenses may be available for both. In particular, one can take a standard Dutch book argument for reflection and modify it to give a Dutch book for conglomerability. However, as I pointed out above, there are serious worries about the weight that should be given to Dutch book arguments, and especially to ones that involve infinitely many bets (as any Dutch book for conglomerability will, except for ones supporting the special case that follows from the standard probability axioms).

Instead, the primary argument I have in favor of conglomerability is a more purely conceptual argument based on Bayesian confirmation theory, as discussed in section 3.3.2. If degree of belief underlies the best means we have for understanding the notions of confirmation, evidence, and justification, then facts about these notions may put some constraints on what sorts of probability functions are acceptable for representing rational degrees of belief.[5]

Of course, in order for this sort of constraint to be convincing, we need to be sure that we properly understand just what notion confirmation theory is discussing. [Maher, 1996] points out that the notion analyzed by Bayesian confirmation theory is not evidence as standardly understood. Standardly, evidence is information that one already has—but this just means that Glymour's "problem of old evidence" always arises. Maher considers some attempts to resolve this problem and shows further problems with each attempted resolution. However, I think this discussion is somewhat irrelevant—I will concede that the problem of old evidence indicates that the notion of confirmation employed by Bayesian confirmation theorists can't be the standard notion of evidence. Instead, the concept under consideration must be some sort of prospective one—to say that E confirms H is to take E to be a *potential* piece of information that would be evidence for H.[6]

[5]Even if there is a good argument that *no* confirmation notion depends on degree of belief, this doesn't mean that my arguments based on confirmation theory won't apply. After all, as mentioned in section 3.1.2, there are connections between degree of belief and whatever the relevant evidential or logical notion is. I pointed out there that there is no necessary connection between their formalisms—norms for one may not be norms for the other, even if degrees of belief ought to track evidential probabilities. However, there is still a *prima facie* suggestion that they should obey the same norms. It is possible that one notion should obey a norm, and the other notion should violate it, but this would require a specific argument. My arguments here will suggest that at least one of these notions (most likely the degree of belief notion) should obey the principle of conglomerability, and I rebut the arguments that degrees of belief should not obey it. Thus, the burden of proof will be on the opponent of conglomerability for degrees of belief to provide new arguments against it, even if the relevant notion for confirmation is a logical or evidential probability.

[6]This distinction alone doesn't fully resolve the problem—we must still for at lest some cases take the resolution offered by [Garber, 1983]. To properly make sense of the historical relation between the well-known perihelion shift of Mercury, and Einstein's theory of general relativity, we must use the notion of prospective confirmation. When Einstein set out to calculate what his theory said about the perihelion of Mercury, he must have known ahead of time that a calculation giving one particular value would strongly confirm his theory, while calculations yielding other values would strongly disconfirm his theory. We must somehow give up on logical omniscience in order to make sense of this prospective confirmation, even once we concede that the notion of confirmation we are interested in is merely prospective, rather than the ordinary notion of evidence that one already has. Of course, much more work must also be done to relate this notion of prospective confirmation to the ordinary notion of having evidence for something—perhaps the second half of [Maher, 1996] can be seen as giving a theory of this other notion, and showing how it is and isn't connected to the prospective notion I am primarily interested in here. See also the distinction between "confirmation" and "evidence" in [Eells and Fitelson, 2000, p. 669]

But even when thinking of confirmation in this more technical sense, there are some clear intuitions about it. In particular, it seems clear that on this prospective notion of confirmation, one should not be able to see every possible outcome of an experiment as confirmatory, and similarly, one should not be able to see every possible outcome as disconfirmatory. If confirmation of a hypothesis H by evidence E occurs only if $P(H|E) \geq P(H)$ and disconfirmation occurs only if $P(H|E) \leq P(H)$, then this requirement just is the principle of conglomerability.[7] To see this, just note that when one has set up some experiment, the possible outcomes of the experiment induce a partition on the space of epistemic possibilities. If every possible outcome were disconfirmatory, then this would mean that for each $E \in \mathcal{E}$, we would have $P(H) > P(H|E)$. But this entails that $P(H) \geq \sup_{E \in \mathcal{E}} P(H|E)$, with equality only possible if the supremum is not actually achieved. Thus, conglomerability for a particular partition just means that not every element of the partition is disconfirmatory. Further work is needed to justify the claim that conglomerability should hold for *every* partition, since this only establishes it for partitions that correspond to experiments. However, these are the only partitions that my argument in the following chapter will require.

One might also consider a slightly stronger principle that also seems to be true. It is not permissible to see one possible outcome of an experiment as confirmatory unless some other possible outcome is seen as disconfirmatory. (The previous version only required that at least one possible outcome is neutral, rather than requiring that at least one be disconfirmatory.) This translates into the slightly stronger requirement that $P(H) \leq \sup_{E \in \mathcal{E}} P(H|E)$, with equality only permissible if $P(H|E)$ is in fact constant for $E \in \mathcal{E}$.

Roy Sorensen has raised the objection that this requirement on confirmation means that no experiment can confirm the claim that at least one object exists. After all, no conceivable outcome of an experiment can disconfirm this claim. However, this seems right to me—no experiment can confirm the claim that at least one object exists because I am already subjectively certain of this claim. Perhaps there can in fact be evidence for this claim (especially since I seem to know this claim, and one might think that knowledge requires

[7]As pointed out in [Fitelson, 1999], many arguments in Bayesian confirmation theory depend closely on which precise measure of confirmational strength is used. However, this particular requirement is common to most measures that have been seriously considered, and in particular to all measures considered there. [Christensen, 1999] argues that in fact there is no unique measure that ought to be used, but even on this pluralistic account, the fact that confirmation occurs only if $P(H|E) \geq P(H)$, and that disconfirmation occurs only if $P(H|E) \leq P(H)$ is common to all permissible measures. The reason for using $\geq$ and $\leq$ instead of the more obvious $>$ and $<$ is discussed in footnote 8.

evidence), but this is just to point out further the fact that the notion of confirmation discussed here is prospective, and not the same as the notion of evidence.[8] It's also important to note that the notion of confirmation being discussed here is always a notion of the type Carnap called "increase in firmness" rather than "firmness"—although it's true that the claim that at least one object exists is always very highly confirmed in the firmness sense after any observation, it seems that no observation leads to an *increase* in this confirmation.

7.3.1 Arguments against conglomerability, and why they fail

It turns out that conglomerability is quite a strong principle. As mentioned earlier, and shown in Appendix A, conglomerability is equivalent to the following equation. Let $\mathcal{E}$ be some partition of a probability space into disjoint events E_α. Let $f_A(w) = P(A|E_\alpha)$, where E_α is the unique element of the partition containing the point w. Then for any B that is the union of some collection of E_α, $P(A\&B) = \int_B f_A(w)dw$. Surprisingly, this argument about the value of an integral doesn't explicitly use countable additivity at any point.

However, as shown in [Schervish et al., 1984], the full version of conglomerability *entails* countable additivity. In fact, as discussed in [Hill and Lane, 1986], if the underlying space of possibilities is countable, then full conglomerability is *equivalent* to countable additivity of all conditional and unconditional probabilities. Thus, reasons to doubt countable additivity are reasons to doubt conglomerability. However, the arguments around countable additivity seem stalemated, so conglomerability may be a better principle to argue for or against.

Several arguments against conglomerability are presented in [Kadane et al., 1986]. Beyond the arguments that rely on finitely (but not countably) additive probability distributions, they suggest an argument based on the decision-theoretic notion of admissibility, and an argument based on the distinction between conglomerability in countable partitions versus all partitions. (They also show several other interesting results related to conglomerability, such as the fact that even if two probability distributions are both conglomerable in a particular partition, a linear combination of them may fail to be conglomerable in that

[8] This also seems the appropriate moment to mention that the criterion of confirmation that $P(H|E) > P(H)$ presents problems when H is not certain, but $P(H) = 1$. I argue in Chapter 5 that this situation should in fact be considered in our theorizing. Since in this situation it seems that any E entailing H should be taken as confirmatory, I only used $P(H|E) \geq P(H)$ as the necessary condition for confirmation. Of course, when $0 < P(H) < 1$, we can safely use strict inequalities instead of weak ones in giving conditions for confirmation, but I use the weak ones to ensure greater generality, and since they suffice for my purposes.

particular partition, and many others.)

Admissibility is the following principle: consider actions A and B, and partition $\mathcal{E}$—if for every $E \in \mathcal{E}$, the expected value of A conditional on E is greater than the expected value of B conditional on E, then B is said to be inadmissible, and so A should be preferred to B *unconditionally*. As pointed out in [Kadane et al., 1986, p. 61], violations of conglomerability automatically give rise to violations of admissibility. Let C be some event such that $P(C) > x$, but such that for all $E \in \mathcal{E}$, $P(C|E) < x$. Then if B is a gamble that pays 1 iff C is true, while A is just a fixed payoff of x, we see that B is inadmissible with respect to this partition. Since admissibility thus entails conglomerability, which itself entails countable additivity, they point out that axiomatizations of decision theory can drop countable additivity as an axiom, if they have admissibility.

Because of the analogy between admissibility and conglomerability, one might suspect that arguments against admissibility should also be arguments against conglomerability.[9] However, [Chalmers, 2002] demonstrates a case where admissibility fails, even though the space is countable and countable additivity holds, and therefore conglomerability holds. (This is not the point of that discussion, which is just to show that admissibility, which underlies the standard St. Petersburg paradox, is an assumption that should be dropped.) As pointed out there, this sort of failure of admissibility can only arise when the unconditional expectation of the actions A and B are both infinite, and thus only when the conditional expectations on elements of E are unbounded. Thus, this sort of problem is avoided if admissibility is only assumed when the conditional expectations are bounded. But this principle of bounded admissibility is sufficient to entail conglomerability, because of the example in the previous paragraph. Thus, even though general admissibility must fail, a very natural restriction of it to bounded cases is sufficient to entail conglomerability, and thus admissibility and conglomerability do not stand or fall together.

Considering uncountable partitions gives perhaps stronger arguments against conglomerability. [Kadane et al., 1986, p. 70] shows that even though countable additivity entails conglomerability in countable partitions, it is still possible for conglomerability to fail in an uncountable partition for a countably additive distribution. In fact, they go on to show that for a certain unconditional distribution (which seems like quite a natural

[9] The authors in fact do not make this argument—they merely point out that if merely finitely additive distributions are allowed, then admissibility should be given up. But this suggests an argument against conglomerability nonetheless.

one, being just a joint distribution of two independent random variables with a normal distribution), *any* way of filling in conditional probabilities gives rise to an instance of non-conglomerability in some uncountable partition. My response to this sort of example is to suggest (as I will do more clearly in the next chapter) that the conditional probability of A given B must be defined only relative to a partition $\mathcal{E}$ containing B as $P(A|B, \mathcal{E})$, and not absolutely by some function $P(A|B)$. I will not discuss the example they consider, but instead a somewhat simpler one. As I will point out there, every unconditional distribution is compatible with relativized conditional distributions satisfying conglomerability in all partitions, and thus this sort of argument cannot rule out full conglomerability.

A further argument against conglomerability is offered in [Arntzenius et al., 2004, pp. 257-260]. This argument relies on two examples in which it looks like the intuitively natural credences are non-conglomerable in some partition. However, the argument can be resisted either if the cases should be deemed epistemically impossible, or if the apparently natural credences can be resisted.

Both examples rely on a dart being thrown at a unit square board, where the agent's unconditional credence distribution for where the dart hits the board is uniform. The first example follows a construction of Sierpinski (which assumes the Axiom of Choice and the Continuum Hypothesis) on which A is a subset of the dart board such that every horizontal slice of A has Lebesgue measure 1 and every vertical slice of A has Lebesgue measure 0. (A itself has no well-defined Lebesgue measure.) Because of these facts, it looks like a partition of the board into horizontal slices must be such that the probability of A given any slice must be 1, so the conglomerability entails that the unconditional probability of A must be 1. A similar argument using the partition into vertical slices suggests that the unconditional probability of A must be 0, which gives a contradiction.

The second example is more straightforward. The authors suggest that there must be some $\epsilon < 1$ such that for any line ℓ and any point p on the board, the probability $P(p|p \vee \ell) < \epsilon$. Given this ϵ, we find some region A, which is a vertical rectangular region of the board with width less than $1 - \epsilon$, and height equal to 1. By some straightforward facts of cardinality (which don't depend on the Axiom of Choice or any other controversial set-theoretic principle), there is a one-to-one pairing of the vertical lines in A with the points not in A. The unions $p \cup \ell$ of corresponding pairs of point and line give a partition of the entire board. However, $P(A) < 1 - \epsilon$, while for any pair of corresponding p and ℓ, $P(A|p \vee \ell) = P(\ell|p \vee \ell) > 1 - \epsilon$, and thus there appears to be a violation of conglomerability.

The natural response to the first example is to point out what I already discussed in section 4.2.1—since degrees of belief are mental states of finite agents, they only apply to events that can actually be considered by the agent. Since the Sierpinski construction requires both the Axiom of Choice and the Continuum Hypothesis, the relevant set can't even be entertained by an agent as a potential object of degree of belief. Thus, if the example is read in this way, then there isn't even a question of what the conditional probability should be, since the agent can't even form this question.

However, Arntzenius, Elga, and Hawthorne present this example in a way that avoids that worry—the relevant set A is constructed by some external being, who also may be assumed to be a reliable informant who tells the agent that the demonstrated set really does have the stated properties. However, if this is the set-up, then the space of possibilities is not properly identified with the set of points on the board. Although it may be true that the possibilities for the location of the dart are actually exhausted by the points on the board, these do not represent the epistemic possibilities for the agent. Since the agent doesn't know which points are in A and which are not (she may know this information for some points, but certainly not for anything like all of the points), introducing this set A has changed the epistemic space for the agent. In addition to the location of the dart, she is now also uncertain of whether each point is in A or not. The new space of possibilities is much more complicated than the original space, but in some ways it's easier to assign probabilities. Whichever point the dart hits (except possibly for points on the diagonal, which are somewhat special in Sierpinski's construction), there are two possibilities for whether the point is in A or not. If the agent is able to use some sort of principle of indifference here, then she can assign probability 1/2 to the dart hitting A. Further support for this assignment comes from the fact that whichever set A actually is, the details of the construction tell us that every point off the diagonal is in A iff its mirror image in the diagonal is not. Since all these facts are completely independent of which line the dart hits, it looks like the agent is justified in assigning probability 1/2 to the dart hitting A conditional on any horizontal or vertical line, thus disrupting the argument. (It is not essential to my point that 1/2 be the appropriate value—I just need the range of values conditional on horizontal lines to overlap the range of values conditional on vertical lines.)

Note that this resolution of the problem requires a distinction between the extensional notion of whether or not the dart actually hits A, and the intensional notion. On any vertical line ℓ, there are only countably many points in A. If this set of points happens to be definable (which may well be the case for many lines), then the agent can consider the set B of points on the horizontal lines through this countable set. Then, since $A \cap \ell = B \cap \ell$, one might suspect (following Section 7.5 below) that the agent should assign equal degree of belief to the dart hitting A and B conditional on it hitting ℓ. However, since the agent doesn't know these two conditions are relevantly coextensive, the events of hitting A and ℓ, and of hitting B and ℓ are *different* events in her possibility space. Thus, they don't have to have the same conditional probability.

The other example looks at first more problematic—the relevant sets and partitions are all definable, so I can't as easily just reject the case, or require that the space of possibilities be considered differently. However, since the account of conditional probability that I will endorse in the next chapter defines conditional probability only relative to a partition, I can point out that the strangeness of any bullet-biting response I make is due entirely to the strangeness of the partition in question. The particular claim I must endorse is that for any $\epsilon < 1$ and vertical rectangular region A of width at most $1 - \epsilon$, there must be some p, ℓ pair in the relevant partition such that $P(p|p \cup \ell) \geq \epsilon$. In fact, there is a sense in which most pairs must be like this, so that the integral works out properly. However, this is perfectly consistent, and is just a result of having conditionalized on a very strange partition. Thus, even the apparent counterexamples to conglomerability described in [Arntzenius et al., 2004] aren't relevant, and thus the arguments in its favor must prevail.

7.4 Symmetry Preservation

A further type of constraint on degrees of belief that is often proposed is the Principle of Indifference (or Principle of Insufficient Reason), which says roughly that an agent's degrees of belief in propositions should be equal, if she has no evidence in favor of one over the other. Some version or other of this principle underlies most "objective Bayesian" accounts of degree of belief (see for instance [Jaynes, 2003] and his "Principle of Maximum Entropy"). Many problems have been posed for versions of this theory, notably including Bertrand's paradox of the chord, which uses versions of the Indifference Principle to produce three contradictory values for the probability that an unknown chord in a circle is at least $\sqrt{3}/2$ times as long as the diameter. These problems lead many to reject the general principle, though [Jaynes, 1973] points out that there are many instances of these principles that are generally considered unproblematic, so a better response is to find a suitably weakened form of the principle that can be accepted.

Following [Jaynes, 1973] and [Bartha and Johns, 2001], I will suggest that the best way to understand this sort of principle is as requiring a certain sort of invariance of probabilities under symmetries of the space of epistemic possibilities.[10] [Bartha and Johns, 2001, p. S116] points out that they do not give necessary and sufficient conditions for a transformation of the epistemic space to count as a symmetry—I will follow them in this regard, and just give necessary conditions, though I hope that the examples I discuss will all be such that they clearly count as symmetries. Where these other authors seek to argue that agents are rationally required to obey a principle of invariance under symmetries, I don't need this very strong claim. For my examples, I just need the claim that if a given unconditional probability distribution is permissible, then its extensions to conditional distributions that obey the symmetry conditions are always permissible as well. This weakening of the principle both allows it to be compatible with some very subjectivist Bayesian pictures (rather than requiring objective Bayesianism) and also makes it easier to argue for in many cases. Bartha and Johns argue that the symmetries must be respected because the agent literally has no way to differentiate an event in the possibility space from its transformation under a symmetry—thus she must assign equal probabilities (and conditional probabilities) to both. However, I allow that the agent may be able to single out particular states or events from a symmetry class, and yet it is still permissible for her to assign equal conditional and unconditional probabilities to all of them.

A symmetry, as I define it here, is some permutation of the agent's space of epistemic possibilities. That is, (following Chapter 4) if Ω is the set of possibilities, then a symmetry is a function $\sigma\colon \Omega \to \Omega$ such that each $\omega \in \Omega$ has a unique ω' such that $\omega = \sigma(\omega')$. Because of this uniqueness and existence (meaning that σ is one-to-one and onto), the inverse σ^{-1} is also a permutation of Ω. Additionally, I require that a symmetry and its inverse both be such that they send events to events—that is, they are measurable functions. Formally, this means that if $\mathcal{A}$ is the agent's algebra of events, then for any $A \in \mathcal{A}$, if we let $\sigma(A) = \{\sigma(\omega) \colon \omega \in A\}$, we have both $\sigma(A) \in \mathcal{A}$ and $\sigma^{-1}(A) \in \mathcal{A}$. (Since $\mathcal{A}$ is the collection of all subsets of Ω that have probabilities, this is necessary for imposing connections between $P(A)$ and $P(\sigma(A))$.) Because I (unlike Jaynes, or Bartha

[10] [Strevens, 1998] points out that there is a difference between the project of inferring physical probabilities from known physical symmetries of objects, and the project of relating subjective probabilities to symmetries of the subject's space of epistemic possibilities, which is determined by ignorance. I am interested here only in the latter project, though the former is surely quite important as well.

and Johns) am primarily interested in conditional probabilities here, rather than unconditional probabilities, I will also say that a necessary condition for σ to be a symmetry is that $P(A) = P(\sigma(A))$ for all $A \in \mathcal{A}$.[11] Finally, it seems appropriate to require that the set of all symmetries forms a group—that is, the identity e such that $e(\omega) = \omega$ for all $\omega \in \Omega$ is a symmetry; for any symmetry σ, its inverse σ^{-1} is a symmetry; and for any two symmetries σ and τ, their composition $\sigma \circ \tau$ given by $(\sigma \circ \tau)(\omega) = \sigma(\tau(\omega))$ is a symmetry.

It's not clear what more may be required to be a symmetry, but these conditions are clearly not sufficient. For instance, if the agent is uniformly uncertain about the location of a point on the surface of a sphere, then the group of symmetries contains all rotations of the sphere. However, some permutations that should clearly not count as symmetries could also be included, consistently with all the above requirements. In particular, if we consider the permutation that leaves every point not on the equator fixed, and does some obviously asymmetrical permutation on the equator (say, compressing every interval containing a specific point towards that point) then it satisfies the above conditions, and yet we intuitively don't want to count it as a symmetry. Some obvious further conditions that would rule out this permutation don't seem to generalize to epistemic spaces in general—for instance, we might want to require that a symmetry be a continuous function on the surface, or that it preserve some notion of measure that distinguishes among events of probability 0. Thus, it's not clear how to complete the analysis of the notion of symmetry other than appealing to intuitions for which transformations should count—however, it is important to make the above conditions explicit.

Given this notion of the group of symmetries of an agent's epistemic space, I can now formulate a few more notions and then state the principle I will endorse. A *subgroup* H of this group G is just a set of symmetries that contains inverses and compositions of any symmetries in H. I will say that an event A is *fixed* under a symmetry σ iff $\sigma(A) = A$—note that this does not require that $\sigma(\omega) = \omega$ for any $\omega \in A$. I will also say that A is fixed by a subgroup H iff A is fixed by every $\sigma \in H$. If $\mathcal{E}$ is a partition of Ω (that is, it is a subset of $\mathcal{A}$ so that every $\omega \in \Omega$ is in exactly one $E \in \mathcal{E}$), then I will say that $\mathcal{E}$ is *preserved* by σ iff for each $E \in \mathcal{E}$, $\sigma(E)$ is also in $\mathcal{E}$. Similarly, I will say that $\mathcal{E}$ is preserved by a group

[11]The other authors seek to understand the relevant group of symmetries first, and impose restrictions on the unconditional probability distribution based on these. Since I don't want to put too much weight on the notion of symmetries, I am allowing for a more pragmatic understanding of them, and thus just want to make sure that whichever permutations are counted as symmetries, they should only be ones that already obey the relevant connections among the unconditional probabilities.

G of symmetries iff it is preserved by every $\sigma \in G$. Now I can finally state the restriction on conditional probability. (Whether it is *really* a requirement of any coherent notion of subjective probability will most likely depend on the account that is finally given of what it is to be a symmetry.)

> Let G be some subgroup of the group of symmetries of the space, let A be a set that is fixed by G, and let $\mathcal{E}$ be a partition of the space that is preserved by G. Then it is permissible[12] that $P(A|E_i) = P(A|\sigma(E_i))$ for every $\sigma \in G$ and $E_i \in \mathcal{E}$.

If G acts transitively on $\mathcal{E}$ (that is, if every event E_i is sent to each other event E_j by some $\sigma \in G$) then this means that A is "independent" of the partition $\mathcal{E}$ in one of the senses mentioned above, in that the conditional probability of A on any element of $\mathcal{E}$ doesn't depend on which element is chosen. Requiring this principle only for transitive group actions on a partition is a way to weaken the principle.

7.5 One Last Principle

One final principle I will require is the following:

> If $A \cap E = B \cap E$ then $P(A|E) = P(B|E)$.

When $P(E)$ is non-zero, this is clearly implied by the standard axioms of probability. One argument for this principle is to assume that one's degrees of belief conditional on E themselves form a probability function. (This is extremely plausible, since any violation of this condition would give rise to a situation in which updating by conditionalization makes one's credences no longer probabilistic.) If we denote this function by P' (so that $P'(X) = P(X|E)$), then additivity entails that $P'(A)+P'(B\wedge\neg A) = P'(A\vee B) = P'(A\wedge\neg B)+P'(B)$. Since $A \cap E = B \cap E$, the agent believes that $B \wedge \neg A$ and $A \wedge \neg B$ both entail $\neg E$. If $P'(\neg E) = 0$ (as seems quite reasonable, though I will mention in the following chapter some potential problems for this principle) then this means that $P'(B \wedge \neg A)$ and $P'(A \wedge \neg B)$ should also be 0, and thus we have $P'(A) = P'(B)$, as claimed.

A more direct argument for this principle is to consider the meaning of the relevant conditional probabilities. Since these are credences restricted to the set of possibilities where

[12] Or perhaps required, if one wants to endorse a stronger version of the principle, like that of [Bartha and Johns, 2001].

E is true, this means that two events that are coextensive within E (as A and B are) should have the same conditional credence, given E. This can be supplemented with a Dutch book argument of whatever sort one endorses for conditional probabilities in general, since a bet on A given E, and a bet on B given E have exactly the same payoff conditions.

Chapter 8

Relativization

In this chapter I will give my first argument that conditional probability can not in general be taken as the fundamental doxastic notion. I will claim that conditional and unconditional probability must both be taken as basic. But in particular, the value of $P(A|B)$ will depend on a set of alternatives to B that are somehow presented as salient—thus, it will be better to think of the notion as $P(A|B, \mathcal{E})$ for some partition $\mathcal{E}$ of Ω, rather than just thinking of it as $P(A|B)$. I will argue for this conclusion using the principles outlined in the previous chapter, and then show why it is not so surprising that the set of alternatives for B comes into the picture somehow. In particular, I will show that in a relatively simple space, there is no assignment of conditional probabilities compatible with the natural unconditional probabilities, given these principles.

Of course, it will be an available alternative to reject one or another of the principles mentioned above. However, I aim to show that this relativized notion of conditional probability is no worse than a non-conglomerable or non-symmetric probability. If there are clear problems with non-conglomerability and non-symmetry, and no significant problems for the sort of relativization I discuss, then we should accept the relativization.

The particular example I have in mind involves an unknown point on the surface of a sphere, where the agent is uniformly uncertain about the location of the point.[1] Thus, the unconditional probability of the point being in any region is proportional to the surface area of that region, and in particular, the probability of any line or point is 0 (or at least miniscule). For simplicity's sake, I will identify the state space Ω with the surface of the

[1] This example is a classical one. It is discussed in [Kolmogorov, 1950] as the "Borel Paradox", but it actually goes back to [Bertrand, 1889].

sphere.[2]

Fix an axis of this sphere, and let A be a region (not the whole sphere) consisting of two congruent discs centered on opposite ends of this axis. (See Figure 8.1.) Let $\mathcal{E}$ be the collection of all great circles on the sphere going through this axis. [3] Let G be the group of all rotations about this axis. Symmetry preservation entails that $P(A|E) = P(A|E')$ for any $E, E' \in \mathcal{E}$. Together with the weak form of conglomerability, this entails that $P(A|E) = P(A)$ for any $E \in \mathcal{E}$.

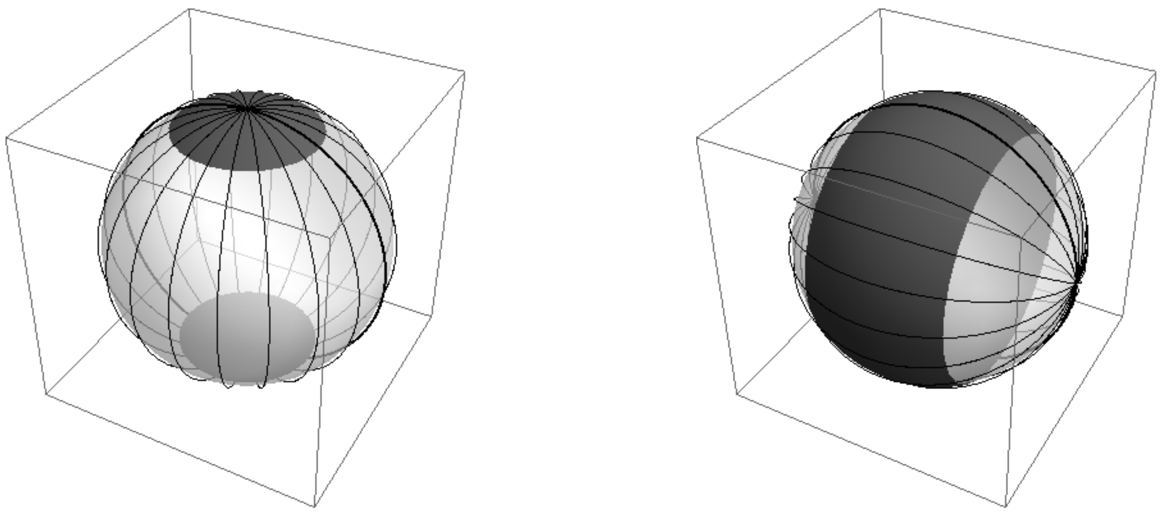

Figure 8.1: Left: A and $\mathcal{E}$. Right: A' and $\mathcal{E}'$.

Now consider some axis perpendicular to this one, and let G' be the group of rotations around that axis. Let A' be the union of the set of all images of A under operations

[2]The caveat mentioned in response to the Arntzenius, Elga, and Hawthorne example in the previous chapter won't be relevant here. The only thing the agent will be uncertain of here is the position of a point on the surface of the sphere, and thus the state space can be identified with this surface. There are no regions whose extent the agent is uncertain of, so there are no extra relevant epistemic possibilities.

[3]Technically speaking, $\mathcal{E}$ is not a partition of the sphere, because the two endpoints of the axis appear in every element of $\mathcal{E}$. One way to fix this problem is to remove these two points from the great circles, and add them as a further element of the partition and use the slightly generalized version of the symmetry preservation principle from the previous chapter, rather than the version where every element of the partition can be sent to every other element under some symmetry.

Another way to deal with the problem that might be simpler is to just remove those two points from the space entirely—then the remaining parts of the great circles really do form a partition of the remaining part of the sphere. Of course, both of these modifications will require further extension when dealing with the second partition that I will describe further on.

Alternately, one can just note that there are only two points that cause problems, while all the other regions in question (both the discs and the circles) has uncountably many points. Thus, intuitively, adding or removing these two points should make no difference to any of the relevant conditional probabilities, so we should just be able to ignore the problem.

of G', which is just the band traced out by rotating A around the new axis. Let $\mathcal{E}'$ be the set of great circles through the new axis. Now, by the same symmetry and conglomerability arguments, $P(A'|E)$ should equal $P(A')$ for any $E \in \mathcal{E}'$.

But now let E be the unique great circle through both axes. By the definition of A', $A \wedge E$ occurs iff $A' \wedge E$ does. By the last principle of the previous chapter, we see that $P(A|E) = P(A'|E)$. But since $P(A) = P(A|E)$ and $P(A'|E) = P(A')$, this means that $P(A) = P(A')$, which is clearly false.

At this point there are several options available. One can deny the cogency of the example (perhaps trying to use some argument like that given in section 4.2.1), or in some other way argue that this example should not play a role in our theorizing. Or one can reject one of the principles used in the argument—symmetry preservation, conglomerability, and the last principle from the previous chapter. Or finally, as I endorse, one can suggest that conditional probabilities only make sense relative to a partition, rather than just as a function of two events—the first symmetry argument will only apply to $P(A|E)$ relative to the partition using G, while the second one will only apply to $P(A'|E)$ relative to the partition using G', and thus even replacing A by A' leaves the partitions distinct, and thus there is no contradiction.

Rejecting the example is fairly radical. As I argued in Chapter 5, there are plenty of examples where scientists are interested in probability spaces that can be parametrized by real numbers. If multiple real numbers are involved, situations similar to the one I discuss here can easily be constructed, by using partitions given by different types of formulas relating the parameters. Unless one is prepared to reject this general type of probability space *a priori*, some version of this situation will be a possibility. (A somewhat more complicated, but only somewhat less radical approach is discussed in Chapter 9.)

Rejecting one of the principles used is also somewhat radical. I have given what appear to be strong arguments in favor of conglomerability, symmetry preservation, and the last principle in the previous chapter. The last principle is exceedingly innocuous, and very well supported, so it doesn't seem like a candidate for rejection. As it turns out, symmetry preservation doesn't appear to be necessary for the argument. As I discuss in Appendix B, there is a construction that satisfies conglomerability and the last principle while avoiding relativization. However, it appears that any such construction requires the Axiom of Choice, and therefore is undefinable.[4] As I argued in section 4.2.1, probabilities

[4]In some ways this is just a weaker form of the claim in [Bartha and Johns, 2001] that agents must respect

are the mental states of finite agents, and therefore can't require the Axiom of Choice for their definition. Thus, although it is mathematically consistent to have such a construction, this can't occur in any agent's mental states. Thus, giving up symmetry preservation won't work.

Thus, the only alternative to the relativization is to give up conglomerability. However, since conglomerability is essential for some basic platitudes about confirmation theory, this is also not a very attractive option. We either have to separate confirmation theory from degree of belief (though see the issues I discuss in section 3.3.2), or accept relativization. In the rest of this chapter, I will argue that the latter option is better, in terms of being less damaging to the familiar theory of subjective probability.

8.1 Relativism is not a problem

My specific proposal is that conditional probability should be a function $P(A|B, \mathcal{E})$, where $\mathcal{E}$ is a partition of the epistemic possibilities, and B is a member of $\mathcal{E}$. This is very similar to a proposal that is quite common in the mathematical literature on measure theory (it is already discussed in [Kolmogorov, 1950, pp. 50-51]), which makes conditional probability a function $P(A|B, \mathcal{B})$, where $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{A}$, and B is a member of $\mathcal{B}$. My proposal can be related to this one by defining $\mathcal{B}$ to be the algebra of all sets in $\mathcal{A}$ that are unions of sets in $\mathcal{E}$. Thus, it is the maximal sub-σ-algebra of $\mathcal{A}$ that contains no elements that cross-cut the elements of $\mathcal{E}$. If the original σ-algebra corresponded to the events we could concern ourselves with, then by taking the maximal sub-σ-algebra with these atoms, we make sure to include every event we could directly learn about by means of the partition. My proposal is a special case of this general one, because of two important facts. First of all, not every sub-σ-algebra of a given σ-algebra is given by a partition in this way.[5] And secondly, not every partition will be epistemically relevant.[6] But since my

symmetries. They claim that if there is an actual complete symmetry to some structure, then the agent can't have different relations to different parts of it, and therefore must assign degrees of belief symmetrically. I'm just saying that certain asymmetric distributions are so asymmetric that no facts about the agent could possibly serve to define the values of the distribution, even if there are facts that break the symmetry in some weaker way.

[5]For instance, if $\mathcal{A}$ is the complete σ-algebra of all sets of real numbers, and $\mathcal{B}$ is the σ-algebra consisting of all sets of reals that are either countable or have countable complement, then we can see that the only partition that is not cross-cut by $\mathcal{B}$ is the partition of $\mathbb{R}$ into the singletons, but the sub-σ-algebra generated by the partition into singletons is just $\mathcal{A}$ itself.

[6]As before, if the partition is some gerrymandered collection of sets that is undefinable (for instance, if its existence is only provable with the Axiom of Choice), then this partition clearly can't be relevant for

proposal is a special case of this other one, I can appeal to theorems by others showing that for any probability triple $(\Omega, \mathcal{A}, P)$, and for any event A and sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$, there is a way to define $P(A|B, \mathcal{B})$ in such a way that for every $C \in \mathcal{B}$ with $P(C) > 0$, we have $P(A|C) = \int_C P(A|B)dP(B)$ (this is the integral equation of Appendix A). Thus, this means that for every A and every partition $\mathcal{E}$, there is a way to define the conditional probabilities $P(A|E, \mathcal{E})$ on all $E \in \mathcal{E}$ such that conglomerability is satisfied. ([Seidenfeld et al., 2001, Ex. 2.2] is an example of a case in which fixing E and $\mathcal{E}$ gives rise to a function that must violate the probability axioms—however, their Corollary 1 shows that this example makes essential use of a non-measurable set, and is therefore irrelevant for degrees of belief.)

The most natural objection to this sort of proposal is just to insist that conditional probability is not a relativized notion—it must be defined absolutely. "This approach is unacceptable from the point of view of the statistician who, when given the information that [some event] has occurred, must determine the conditional distribution of X_2." [Kadane et al., 1986, p. 70] I have several responses to this point.

First of all, this relativism only arises in the case where we conditionalize on an event of probability 0. In any other case, the constraint that $P(A \wedge B) = P(A|B, \mathcal{E})P(B)$ shows that $P(A|B, \mathcal{E})$ doesn't depend on $\mathcal{E}$. Even in cases where $P(B) = 0$, there may be many A for which $\mathcal{E}$ doesn't matter in determining the conditional probability. For instance, in the Borel paradox case described above, if A is a complete hemisphere and B is a great circle on the sphere, then similar symmetry arguments to those given above suggest that $P(A|B, \mathcal{E}) = 1/2$ when $\mathcal{E}$ is any partition of the sphere into lines of longitude about a given axis. Thus, it's not always necessary to find a specific partition to be able to calculate the relevant conditional probability. There may even be very general sorts of cases in which this occurs—for instance, if $\mathcal{E}$ and $\mathcal{E}'$ consist of all the same sets "in the neighborhood of B", but different sets elsewhere, then it seems that all the relevant conditions imposed by the integral equation should be the same, and thus the conditional probability value should be the same. (The hard part is just saying what "in the neighborhood of B" amounts to, especially since probability spaces don't in general come with the sorts of geometric structure that the Borel paradox case does.)

But these considerations only apply in certain special cases—I must still address

finitary agents. Even for partitions that are definable, it may not be the case that any epistemic situation could correspond to the partition, if the relevant sorts of uncertainty must for instance always come with other uncertainties.

the worry.

8.1.1 Where does the partition come from?

In most ordinary applications of conditional probability, we don't seem to be aware of this partition, so I suggest that it must be contextually filled in in some obvious way. I deny that there is one right way in general to specify the partition,[7] but I should say for each particular use of a conditional probability there is an appropriate partition given the context.

As described in Chapter 3, the three major uses of conditional probability I am considering are for updating, for confirmation theory, and for decision theory. If I can show that finding the relevant partition for each of those uses is unproblematic, then the dependence of my account on a partition is not a serious problem.

Updating

Standard update by conditionalization states that when an agent comes to know some proposition B, then her new degree of belief $P'(A)$ should be given by her old conditional degree of belief $P(A|B)$. I say that in situations where this norm applies, it must be given by $P(A|B, \mathcal{E})$ for some partition $\mathcal{E}$. My suggestion here is that just as B gives the content of *what* is learned, $\mathcal{E}$ will give the manner of *how* it is learned. In some cases it is clear how this will work—if the agent has set up some experiment, and observed B as an outcome, then the partition will be formed of the other possible observations she might have made. If she set up a device to measure the mass of a particle, then the partition is into different masses, while if she was measuring the velocity instead, she would have had a different partition. There is a natural sense in which an experiment thus corresponds to a partition, so that if the updating proposition is learned as the result of an experiment, then the experiment that was actually performed gives the relevant partition.

Even when B was not learned by means of some experiment, but rather from testimony (as in Kadane's worry above) something like this can be applied. If B came as the answer to a question, then the partition is given by the alternate answers to the question that the agent considered as epistemic possibilities. Even if it is just the result of

[7] If there were one partition that were always the right one to turn $P(A|B)$ into $P(A|B, \mathcal{E})$, then the partition would not really be a third independent parameter—but it must be an independent parameter in order to deal with the Borel paradox case described above.

an unexpected announcement by someone, we can still consider the partition of epistemic possibilities given by the other things this person could have said. And in fact, in general, [Schaffer, 2007b] argues that *all* knowledge is best conceived of as knowing the answer to a question. If this controversial position is right, then there is no problem—when the agent comes to know B, she comes to know it *as* the answer to some particular question, and this question will give $\mathcal{E}$.

Now in fact, this general sensitivity of updating to the method by which one learned a relevant proposition is familiar from the Monty Hall "paradox". In this situation, there are three doors, and a prize behind one of the doors. When the contestant picks a door, the host (Monty Hall) will choose a door that is not the contestant's, and does not have the prize, and open that door. The paradox is that if the contestant originally chose door 1, and Monty opened door 3, then the contestant should now assign probability 2/3 to the prize being behind door 2, even though her original credence in the prize being behind door 2, given that it's not behind door 3, was only 1/2. In this situation, the update is sensitive to the fact that she knows the procedure by which Monty will choose which information to give her, so she must update differently. (Of course, I don't mean to suggest that this is an example of the partition relativity I am discussing— this is just an analogy. In this situation, the manner in which the information was arrived at can be rolled into the evidence itself, so that conditioning on the proper event gives the right answer. My suggestion is that in some situations, the manner of learning will show up in the partition $\mathcal{E}$ rather than the event B.)

There is a Dutch book argument in [McGee, 1994, pp. 184-6] arguing that rational agents whose betting behavior is related in a natural way to their conditional and unconditional degrees of belief must be representable by a Popper function.[8] This seems to be in conflict with my claim that these degrees of belief should be relative to a partition. However, McGee's argument specifically considers a partition (which he refers to as "$\{b_j : j \in J\}$") consisting of "every possible course of experience the agent might have between now and time t." Additionally, the only bets McGee considers are bets entered into now, and bets entered into at time t. Thus, for all of these bets, the only conditional probabilities that are relevant are probabilities conditional on the learning experience between now and time

[8] This argument officially seeks to establish that Popper functions and infinitesimals have equal expressive power. However, this argument is quite different in form from the argument to that effect in [Halpern, 2004], because this argument is based on epistemic norms, and is thus in some sense more a positive argument in favor of Popper functions. However, I will show here that it doesn't go quite that far.

t—on my account this means that they are all relative to the partition $\{b_j : j \in J\}$. Thus, McGee's argument actually supports my position on updating—he just expresses his result in terms of a Popper function rather than a relativized conditional distribution because he has fixed a unique partition throughout his entire discussion. Because he assumes (without clear justification) that J is finite, the principles he explicitly mentions entail conglomerability —but if J had been infinite, his Dutch book argument would give conglomerability as an extra condition that must be met. Of course, there are strong worries about Dutch book arguments in general, and this one in particular only addresses the notion of updating one's credences rather than the other things one might do with conditional probabilities, but it still seems to cohere with my approach to conditional probability.

As discussed in section 3.3.1 though, the appropriate way to update is often considered to be by Jeffrey conditionalization, rather than standard conditionalization. (This may be very clear for the example of updating on the result of an experiment—if the researcher can't be completely certain which mass, or velocity, her instruments are showing, then all we can say is that she now has much higher degrees of belief in certain values than in others, and we need the Jeffrey picture to say how this will affect all her other credences.) However, in this case even the traditional picture insists that a partition is relevant. So for Jeffrey updating, there is no new problem.

In fact, Jeffrey updating can even be used to give another positive argument for the picture that I endorse, and for conglomerability in particular. On the traditional picture of Jeffrey updating, there is some partition $\mathcal{E} = \{E_1, \ldots, E_n\}$, and P_{new} is related to P_{old} by the following equation:

$$P_{new}(A) = \sum_{i=1}^{n} P_{old}(A|E_i) P_{new}(E_i).$$

Thus, if $P_{new}(E_i) = P_{old}(E_i)$ for all i, we will also have that $P_{new}(A) = P_{old}(A)$ for all A. (This assumes that there is no E_i such that $P_{old}(E_i) = 0$ but $P_{new}(E_i) \neq 0$.) If $\mathcal{E}$ is a countable partition, rather than finite, then this same strategy can work.

However, when $\mathcal{E}$ is uncountable, this will not work—as described in section 4.2, there are probability spaces that agree on the probabilities for each element of an uncountable partition, and yet disagree on the probabilities of various unions of elements of this partition. Thus, to specify the new distribution it is at least necessary to give a probability distribution for the entire σ-algebra $\mathcal{B}$ that is related to $\mathcal{E}$ as above. (That is, it contains every element of $\mathcal{A}$ that is a union of elements of $\mathcal{E}$.) Once we have this new unconditional distribution, there remains the question of how to generate all the other new probabilities given some old conditional probabilities, and this distribution on $\mathcal{B}$. The natural suggestion is to replace the sum with an integral over this new probability distribution. Finally, we want to make sure that if P_{new} agrees with P_{old} on $\mathcal{B}$, then it should agree on every event in $\mathcal{A}$—but this just amounts to requiring that the *conditional* probabilities on elements of $\mathcal{E}$ satisfy the integral equation of Appendix A. Thus, in order to generalize Jeffrey conditionalization to uncountable partitions, we must have a notion of probability conditional on elements of a partition that is conglomerable in that partition—but this is exactly what I say the notion of conditional probability must be. Once we have these conditional probabilities, it seems that these should be the right ones to plug in to the integral to actually make changes using Jeffrey conditionalization. This still leaves the problem of finding the *conditional* probabilities in the new distribution, but presumably the same method that gave us the old conditional probabilities will give us the new ones as well.

Confirmation

In confirmation theory, conditional probabilities are used in several different ways. However, in general there are two main types of conditional probability that arise for measuring the degree to which evidence E confirms hypothesis H. One is the "posterior probability" $P(H|E)$, and the other is the "likelihood" $P(E|H)$.

For the posteriors $P(H|E)$, the response I gave above for updating seems to work. In general, E is being considered as a potential outcome of some experiment, and this experiment defines the partition $\mathcal{E}$. Thus, $P(H|E)$ can be safely replaced by $P(H|E, \mathcal{E})$, where $\mathcal{E}$ is already understood. This gives us enough information to make sense of the "difference measure" of confirmation $P(H|E) - P(H)$, and the "ratio measure" $P(H|E)/P(H)$, which are two widely used measures.

[Joyce, 1999] and [Christensen, 1999], among others, have argued for a sort of pluralism of confirmation measures, while others have argued in particular in favor of the "likelihood ratio" $P(E|H)/P(E|\neg H)$. Since this makes use of the likelihood $P(E|H)$, I must say where the partition there comes from. In this case, just contextually specifying a partition is unlikely to seem sufficient. After all, the likelihoods $P(E|H)$ are often just identified with the objective chance that theory H says E has. Thus, there must be some uniform partition that is always used to calculate likelihoods. However, such a partition can be given by noting that alternatives to H form a partition over the epistemic possibilities. Alternatives to H do not completely specify every fact—they just specify the laws, or something similar. For instance, H might be the hypothesis that a certain type of radioactive atom has a 50% chance of decaying in any given ten minute period. Then the alternatives here will be the other chances these atoms might have of decaying. Each of these alternatives is compatible with the proposition that ten out of a given hundred atoms of the relevant type decayed in a given ten minute period. Propositions of this sort, specifying particular matters of fact rather than the general chancy laws, may well enter into partitions if H is being considered as a testimonial datum to be learned. But as long as H is being conditioned on for measuring likelihoods, it seems that the correct partition is in terms of the alternative chance theories, rather than any other class of propositions.

Decision

For the conditional probabilities used in decision theory, the problem of finding a relevant partition is quite simple. Since the relevant conditional probabilities are always conditional on the action that the agent is contemplating, the relevant partition of alternatives should be the partition given by the alternate actions the agent is contemplating.

8.2 Further worries

8.2.1 Symmetry

Since the particular argument I gave above makes essential reference to symmetries, it might seem surprising that the specific conditional distributions that result are in some sense very asymmetric. In particular, some calculations will show that on my account, if A is the region consisting of points with latitude between x and y, and if $\mathcal{E}$ is the partition into lines of longitude, and B is some particular line of longitude, then $P(A|B,\mathcal{E}) = \left|\frac{\sin x - \sin y}{2}\right|$. One might expect that since the unconditional distribution is uniform over the surface of the sphere, then the conditional distribution should be uniform over the relevant great circle, B. If G is the group of rotations of the sphere around the axis perpendicular to B, then G fixes B, but different elements of G send A to sets whose probability conditional on B is different from that of A.

However, this should not be surprising. Since the conditional probabilities I discuss are $P(A|B, \mathcal{E})$, rather than just $P(A|B)$, we should not expect the values to be preserved by symmetries acting on A unless those symmetries fix both B and $\mathcal{E}$. Although elements of G all fix B, they change the point on B which is the axis for the partition $\mathcal{E}$. Thus, although G is a group of symmetries of the original epistemic possibility space, these symmetries are broken by fixing $\mathcal{E}$ as a relevant partition. This contrasts with the suggestion on [Bartha and Johns, 2001, p. S114], where they suggest that unconditional probabilities and conditional probabilities may both be generated by the group of symmetries on the space as a whole. I am suggesting that for many uses of conditional probability, there is some relevant partition that is contextually specified, which breaks some of these symmetries, so there cannot be one group of symmetries that generates both the conditional and unconditional probabilities. (Additionally, as I pointed out in the previous chapter when introducing the principle of symmetry preservation, I don't assume that the symmetries are so total that the agent can't even tell the difference between parts of the possibility space that are related by one of the symmetries. Such symmetries seem to me to be quite rare. But if there were such a class of symmetries, then these symmetries could not be broken by any supposition of the agent, or even any experiment she could carry out.)

8.2.2 Propriety

A further worry for my sort of approach to conditional probability is raised in [Seidenfeld et al., 2001]. They discuss "regular conditional distributions", which are the general mathematical functions $P(A|B, \mathcal{B})$, where $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{A}$, and show that in many spaces, such functions must always be "improper". What it means to be improper is that there are non-empty sets $A, B \in \mathcal{B}$, with $B \subseteq A$, and yet $P(A|B, \mathcal{B}) \neq 1$. This would clearly be a problem for conditional degrees of belief, if it were to arise. "If conditioning on a σ-field is to represent coherent degrees of belief, then the rcd should be proper." [Seidenfeld et al., 2001, p. 1616] They show in fact that in some spaces, this problem is so bad that the value can be as low as 0, and that the B's for which this occurs have a union with unconditional probability 1.

However, the examples they give all have strange features. One essentially involves an unmeasurable set; one uses a sub-σ-algebra $\mathcal{B}$ that consists of the countable and co-countable sets (and is thus not generated by any partition); one consists entirely of events containing only countably many states. Thus, it's not clear how relevant their examples are. Indeed, as they point out, [Seidenfeld et al., 2001, p. 1617] whenever the σ-algebra being conditioned on is countably generated[9] (as I suggest it must be for any partition that is to represent the possibilities that a rational agent can consider) the set of points at which impropriety occurs has measure 0. They go on to discuss examples in which impropriety still occurs, but I suspect that these examples too will turn out to be irrelevant for epistemic theorizing.

At any rate, there is still further work to do to determine under what conditions impropriety occurs, and also to determine what sorts of partitions can actually play the relevant role in any of the uses of conditional probability.

8.3 How to Calculate Conditional Probabilities

The integral equation proved in Appendix A guarantees that conditional probabilities are "almost unique", so we may be able to gain some hope of deriving conditional probabilities from unconditional. However, the relevant theorems only prove the existence of functions satisfying this result, and don't give an easy way to find their values. There are also further problems associated with the fact that the function is only *almost* unique. Thus, any way to calculate the value of $P(A|B, \mathcal{E})$ when $P(B) = 0$ will be more complicated than just taking the ratio, as we can do when $P(B) > 0$.

8.3.1 Ways that don't work

One very tempting way to calculate $P(A|B)$, when $P(B) = 0$, is to specify some sequence $B_1, B_2, \ldots$ of events such that $P(B_i) > 0$, $B_{i+1} \subset B_i$, and $\bigcap_{i=1}^{\infty} B_i = B$, and then let $P(A|B) = \lim_{i\to\infty} P(A|B_i)$. The main problem with this method (as discussed in [Hájek, 2003], among other sources) is that it is sensitive to the precise sequence B_i.

For instance, consider a uniform probability distribution on the square $\{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$, and consider the conditional probability $P(y > 1/2|x = 0)$. (See Figure 8.2.) The obvious thought is to take the limit of $P(y > 1/2|x < 1/i)\}$ as i goes to ∞, which gives the expected answer of $1/2$. However, we can also consider the limit

[9] Note that this doesn't require there to be only countably many *atoms*, even if the sub-σ-algebra is atomic. The Borel algebra on $\mathbb{R}$ is generated by the countably many open intervals with rational endpoints, but it has uncountably many atoms.

of $P(y > 1/2|y > ix)$, which effectively lets B_i be a triangle with one side given by the segment of interest, and the other vertex at $(1/i, 1)$, instead of being a rectangle. In this case, each of the conditional probabilities is $3/4$, so the limit will be as well. Thus, we need some way to further specify which sequence of events converging to B is the "right" one, in order to calculate this value.

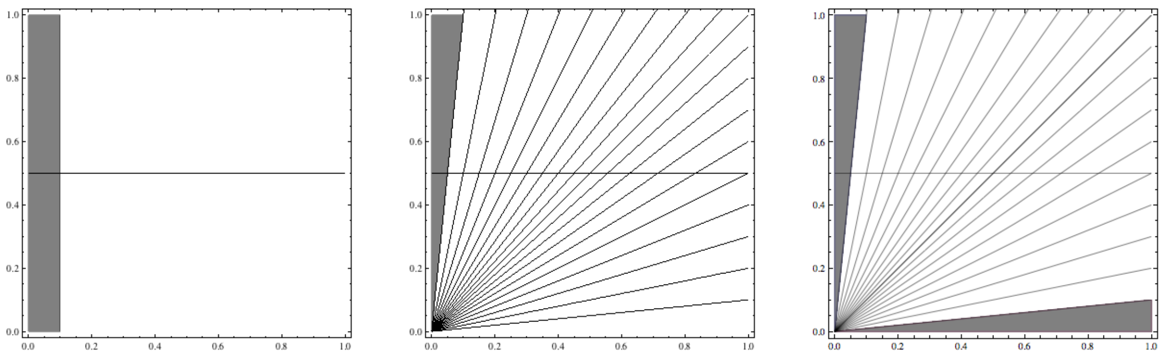

Figure 8.2: Left: $P(y > 1/2|x < 1/i)$. Middle: $P(y > 1/2|y > ix)$, with partition into lines through the origin. Right: $P(y > 1/2|(y > ix) \vee (y \neq 0 \wedge x > iy)$, with partition into lines through the origin.

A more sophisticated version of this method might take into account the partition $\mathcal{E}$—after all, I suggest that this partition can play a role in setting the value of the conditional probability, so it seems that any method of calculating this value must make use of the partition in some way. The natural revision is to suggest that in addition to B being an element of $\mathcal{E}$, we should also require that the B_i are composed of unions of elements of $\mathcal{E}$, so that the sequence is a sequence of events that are relevantly compatible with the partition. Unfortunately, even this won't work. If we return to the example with the square, and let $\mathcal{E}$ be the partition of the square into lines through the origin, then we can still generate two sequences of B_i that converge to B, and yet give different limiting conditional probabilities. The second example discussed above lets B_i be the set of points such that $y > ix$. However, we can also consider the sequence of B'_i given by $(y > ix) \vee (y \neq 0 \wedge x > iy)$. If A is the set of points with $y > 1/2$, then $P(A|B_i) = 3/4$, but $P(A|B'_i) = 3/8$, and thus we get different limiting values. Jake Ross has suggested to me that we might be able to avoid these problems by requiring that the B_i be connected (or perhaps convex) regions. It's possible that imposing some sort of restriction like this may avoid these problems, but this sort of restriction would be inapplicable to general spaces of possibilities, where these topological or geometric notions aren't defined.

Worse still, there are related problems available for how we consider A, and not just B! $P(A|B_i)$ might have a non-zero limit, so this method would say that $P(A|B) \neq 0$. But since $P(A \cap B) = 0$ when $P(B) = 0$, we'll have $P(A \cap B|B_i) = 0$ whenever $P(B_i) > 0$, so the limit method would tell us that $P(A \cap B|B) = 0$. But according to the final principle I endorsed in the previous chapter, $P(A_1|B) = P(A_2|B)$ if $A_1 \cap B = A_2 \cap B$. In particular, this means that $P(A \cap B|B) = P(A|B)$which contradicts the earlier claim.

8.3.2 A way that might work

The one suggestion that does seem to be helpful for calculating the actual values of conditional probabilities is to appeal to the principle of symmetry preservation that I discussed in the previous chapter. I have some worries about this principle in general, because of its connection to the Principle of Indifference, and it also seems that many probability spaces just won't have enough symmetries to permit this sort of calculation, but this method seems to work for the mathematical cases where I have investigated it.

The idea is that symmetry preservation can give a large collection of B's in our partition $\mathcal{E}$ for which $P(A|B, \mathcal{E})$ is constant. Then, we can appeal to one of the versions of conglomerability to argue that this constant value must be equal to some unconditional probability, or at least some conditional probability where we condition on an event of non-zero probability, so that we can actually calculate the value.

More formally, consider some events A and B, together with some partition $\mathcal{E}$ containing B as an element. Assume, as above, that there is some group G of symmetries, and that each element of G sends elements of $\mathcal{E}$ to elements of $\mathcal{E}$. For $X \subset \Omega$ and $g \in G$, let gX be the image of X under g. Let us also assume that $\bigcup_{g \in G} gB$ has positive measure. Then the principle of conglomerability, together with the symmetry principle from above, will require that $P(A|B) = P(A'|B')$, where $A' = \bigcup_{g \in G} g(A \cap B)$ and $B' = \bigcup_{g \in G} gB$.

For a case that seems to generate a problem, but which actually doesn't, let Ω be a Möbius strip, with a uniform probability distribution. (Formally, we can represent the Möbius strip as the set of points in the plane with $0 \leq x \leq 1$ and $0 \leq y \leq 1$, but where the point $(0, y)$ is identified with the point $(1, 1 - y)$, so that we have a strip with the two ends "glued together" with a half-twist.) Let G be the group of rotations along this strip—formally, let g_k be the transformation that sends any point (x, y) to $(x + k, y)$, if $x + k < 1$, and to $(x + k - 1, 1 - y)$ if $x + k > 1$, and let g_{-k} be the inverse of g_k. Let $\mathcal{E}$ be the partition into vertical lines, so that E_i is the set of points (i, y). Note that g_1 and g_{-1} are the same symmetry, and that they both send (x, y) to $(x, 1-y)$. If A is some region that intersects E_i exactly in the upper half of the line, then note that $A' = \bigcup_{g \in G} g(A \cap E_i)$ will actually be the whole Möbius strip Ω. Thus, it looks like the previous calculation will suggest that $P(A|E_i, \mathcal{E}) = P(A') = 1$, which clearly seems wrong. However, this is only an apparent counterexample, because the nature of the example shows that A is not itself fixed under the symmetries in G—if it were, then we wouldn't have the fact that $(A \cap E_i) \cup g_1(A \cap E_i) = E_i$. In this example, the fact that the symmetries in some sense give a "double cover" of the space by itself (for every pair (E_i, E_j) of elements of $\mathcal{E}$, there are exactly *two* elements of G (namely g_{i-j} and g_{1+j-i}) that map E_i to E_j) suggests that we can use a modified version of the previous procedure to calculate conditional probabilities, if we just divide all the resulting values by 2.

In fact, this method can be seen as a fix of the limits described in the previous section. If we just require that A be an event that is fixed under the relevant symmetries, and that the symmetries preserve the partition $\mathcal{E}$, then using *any* sequence B_i that converges to B will give the correct value of the conditional probability, because the symmetry guarantees that all the conditional probabilities will in fact be the same—so taking the limit will be unnecessary.

Another argument by symmetry can be used to calculate conditional probabilities for other partitions. Instead of partitioning the sphere into longitudes, as I did above, partition the sphere into vertical planes parallel to a given great circle. Using symmetries, I can show that in this example, $P(A|B, \mathcal{E})$ really is just given by the fraction of B's length that A covers, as one might have intuitively suspected.

In this case, let A be some a segment consisting of exactly $1/n$ of the whole sphere. (See Figure 8.3.) Let G be the group of n rotations of the sphere which preserve the partition and rotate by some multiple of $180/n$ degrees. Since A is not fixed under these rotations, we can't use the same symmetry preservation principle discussed before. But this time, B is fixed, and so are all the other elements of $\mathcal{E}$. So if we assume that $P(A|B, \mathcal{E})$ should equal $P(gA|gB, g\mathcal{E})$, then this means that $P(A|B, \mathcal{E}) = P(gA|B, \mathcal{E})$, and since B is just the union of the n events gA, this entails that $P(A|B, \mathcal{E}) = 1/n$. Since every measurable subregion of B is generated by these rational half-open segments by countable unions and intersections, the fact that these probabilities are uniformly distributed means that the probability conditional on B with respect to $\mathcal{E}$ should be uniformly distributed as

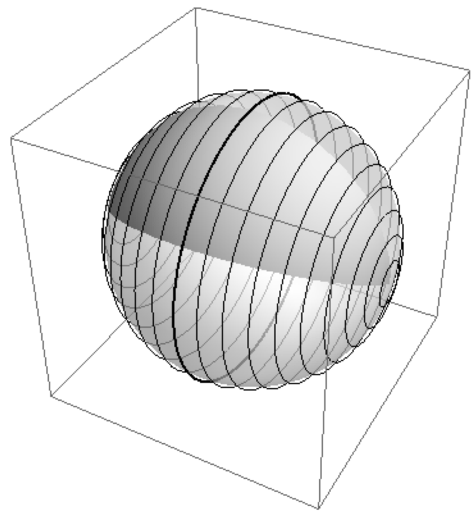

Figure 8.3: Segment A of the sphere, with partition $\mathcal{E}$, and B in bold.

well. Thus, there is a partition that preserves the expected Lebesgue measure, even though it is not the partition that we started with at the beginning of this chapter.

Thus, if this sort of method can be generalized, it looks like I can redeem the basic idea of [Bartha and Johns, 2001]—both unconditional probabilities can be generated by the symmetries of a space. However, because there are so many symmetries of some of these spaces, we have to be careful when choosing which symmetries to look at. Whenever there is a partition of the space that is relevant to the use of the conditional probability (as I have argued that there is when the conditional probability is used for update, confirmation, or decision theory), we must restrict attention to symmetries that somehow respect this partition. But since many epistemic spaces seem to lack any recognizable notion of symmetry, these insights are unfortunately not as general as one might hope.

Chapter 9

Imposing Regularity

A defender of the traditional ratio analysis of conditional probability, might respond to the results of the previous chapter by trying to deny the cogency of conditionalizing on any miniscule event. The first step in such an argument is related to a point that I have made several times (especially in section 4.2.1, which is that not every set of states corresponds to an event. Since events are the only things that appear in probability functions, there are many sets of states that it just doesn't make sense to conditionalize on. Thus, it may be possible to deal with the problem of miniscule denominators by assimilating it to another problem Hájek discusses.

This chapter will outline such a potential defense. I'm not sure whether or not the resulting picture is appealing enough to actually adopt, but it does seem consistent.

9.1 Undefined Denominators

This problem is one that I think the defender of the ratio analysis can deal with relatively straightforwardly. Here is Hájek's description of the problem:

> As a point about probability functions in the abstract, this is familiar enough. We begin with a probability space $(\Omega, \mathcal{F}, P)$, where Ω is some set of 'elementary events, or 'possible worlds, and $\mathcal{F}$ is a σ-algebra of subsets of Ω. P is defined on all the members of $\mathcal{F}$, but on no other propositions. So any subset of Ω that does not belong to $\mathcal{F}$ can be thought of as a probability gap as far as P is concerned: a proposition to which P assigns no value.[Hájek, 2003, p. 278]

Further, when a proposition has no probability, there can be no probability conditional on it—in particular, if X is some such proposition, then $P(X|X)$ comes out undefined on the ratio analysis, even though in some intuitive sense, Hájek suggests that it should be 1. In particular, he considers an unmeasurable subset of the real line, and then considers the event C of an infinitely thin dart thrown at the real line hitting a point within some particular such set. Then he asks what the ratio analysis says about $P(C|C)$.

> You might say that since it fails to yield an answer, there is no answer—that conditional probabilities go undefined when the conditions are non-measurable. This is [a] 'bullet'-biting response. ... But again it strikes me as quite unsatisfactory. When I held it as self-evident that the conditional probability of every (non-empty) proposition, given itself, is 1, no proposition (apart from the empty one) was exempt. Every proposition implies itself, however strange that proposition might be in other respects. C is no exception. The truth of C, *conditional* on the truth of C, is guaranteed.[Hájek, 2003, pp. 300-1]

I will bite the bullet here—fortunately, I think this is not such a bad position.

Let us consider the basic framework again, from Chapter 4—the probability space is $(\Omega, \mathcal{A}, P)$, where Ω represents something like the set of all epistemically possible worlds (states), and P represents the agent's subjective credence function. But we should give more consideration to what $\mathcal{A}$ is supposed to represent.

In one sense, the other stipulations have spelled out that $\mathcal{A}$ is the algebra of events to which the agent assigns credences. So as suggested in section 4.2.1, each element of the algebra should be somehow accessible to the agent, or at least fixed by some feature of the agent's mental state. If we want to make the stronger insistence on accessibility to the agent, then a natural way to phrase this condition is to say that the elements of the algebra are all and only the sets of worlds that are definable in the agent's language, rather than just requiring them to be provable in ZF. (This proposal would presumably have to be modified for non-linguistic agents.) Depending on one's account of semantics, this might mean that they should be exactly the semantic values of the sentences of the agent's language.[1]

[1]Issues may arise here as to whether the bearers of probabilities should be sentences, propositions, meanings of sentences, or something else. These issues are exactly analogous to worries about truth-bearers that arise in other contexts, though there is no clear reason to believe that they must be the same as the truth-bearers. However, it does seem plausible that whatever class of entities we end up picking out as the probability-bearers, they should somehow be distinguishable by the agent, whether by corresponding to distinct sentences of her language, or otherwise. In this chapter I will bracket these concerns and act as though $\mathcal{A}$ is the algebra of sentences (or meanings of sentences) of the agent's language, but parallel points should go through on any other account. Greg Restall pointed out to me that even in a countable language, there might be uncountably many meanings of sentences, because of the distinct contributions of indexicals, demonstratives, and the like. However, if we follow [Titelbaum, forthcoming] and assign credences to *sentences at a time*, rather than something more abstract, then I believe we can avoid this issue. In a countable language, at a fixed time, there will only be countably many such sentences available.

As I have suggested earlier, it is not so problematic to deny the existence of a (subjective) conditional probability $P(X|X)$, when X is not in $\mathcal{A}$. This is because the agent can't even understand the question of what $P(X|X)$ is, because it can't be expressed in her language. She can assent to the claim that for all meanings of meaningful sentences in her language, the probability $P(X|X)$ is 1. However, for an event that's not in her language, it doesn't make sense to ask her to assign a conditional credence any more than it would to ask her to assign an unconditional one. Hájek said "every proposition implies itself, no matter how strange the proposition"—but we may wonder here whether there really is an *event* corresponding to this strange set of states. Besides, questions of logical entailment seem more relevant to the logical interpretation of probability, rather than to the subjective one.

On this account, sets that are non-measurable have no special place among the non-measured sets. Every set that is not describable in the agent's language is equally incapable of having a measure, whether it be some Vitali-style set (which requires the Axiom of Choice to construct, and therefore can't be picked out uniquely by any description) or just some set the agent's language happens not to pick out for some other reason. Thus, there is no problem of conditionalizing on a set without a probability, since the agent can't even conceive of those sets.

Note that in this chapter I am making the strong assumption that the agent does in fact assign a probability to every set describable in her language. I don't mean to presume that she (or anyone else) has access to the values of these probabilities, just that they exist, whether in some categorical or merely dispositional way. Something about her mental state defines her doxastic attitude towards each such set. If not every set in the algebra had a probability, then the problem of dividing by an undefined denominator would arise again. So if there are problems with this assumption, then this resolution of the problem of undefined denominator is incomplete.

Thus, on my account, $\mathcal{A}$ plays a much more important role in the definition of the probability space than one might have assumed. It is not only the algebra of all sets that the agent *does* assign a probability to, but it is also the algebra of all sets that it would *make sense* for the agent to assign a probability to. If every unconditional probability of a set in $\mathcal{A}$ is defined (as Hájek and others seem to presume) then we don't run into the problem of undefined denominators when considering conditional probabilities involving two sets in the algebra. And if one of the sets isn't in the algebra, then it doesn't make sense to ask about the conditional probability anyway, so there is no problem if the ratio analysis doesn't give a value.

Thus, at this point, there is only the problem of miniscule denominators left over.[2] From now on, I will presume that all sets of worlds that an agent is considering are sets in $\mathcal{A}$, and thus will call them "events". I may use the words "state" and "world" interchangeably for the elements of these sets. But as pointed out in Chatper 4, one should not confuse states and events.

9.2 Miniscule Denominators

The problem of miniscule denominators is more tricky. As I discussed in Chapter 7, many authors want to insist on regularity, so that this problem doesn't arise. Unfortunately, as discussed in Chapter 5, in many not-too-unrealistic cases regularity seems pretty implausible.

As a particular example, imagine an agent who believes (with some positive probability p) that a certain coin will be flipped (in a fair way) once a second for the rest of eternity. (Although this belief may seem somewhat crazy, it doesn't seem logically contradictory, so a defender of regularity should require the agent to assign it a positive probability.) Since she knows it is fair, she assigns, for any specification of the values of n of the flips, the credence $p \cdot 2^{-n}$. Now consider the event corresponding to the following sentence, "The coin always lands heads." (The set of states in which this sentence is true is simply defined by this sentence, so as long as the agent's language is not too terribly impoverished, this set is in fact an event, as described above, so she assigns it some credence.) This event H_∞ entails each event H_n of the form "The first n flips of the coin all come up heads." Thus, its probability must be less than each of theirs. But their probabilities are $p \cdot 2^{-n}$, so its probability must be less than all of those. Since probabilities are non-negative, this means the probability must be miniscule.

One response to this example might be to respond that the event defined by this sentence *is* in fact the empty set. As I will point out later, my defender of the ratio analysis does have sympathy for this response, but I suggest that we will have to be very careful.

[2]Hájek also raises a problem of *vague* denominators—I may either suggest that the probabilities aren't vague (we just have imperfect access to them), or else suggest that whatever treatment of vagueness applies elsewhere will generalize to this problem as well.

The standard counter-argument to this response is to point out that not only this sequence, but *every* sequence of specifications for coin flips will be forced (by similar reasoning) to have miniscule probability. So if we take this response to every sequence, then we end up saying that the set of states for *any* complete sequence will be empty. But this contradicts the claim at the beginning that the agent gives positive credence to the claim that the fair coin will be flipped once a second for the rest of eternity!

However, this counter-argument relies on the claim that *every* infinite sequence of coin flips will be forced to have probability zero. But if we consider the framework for subjective probability that I described above, we see that there is a missing presupposition here—namely, that every infinite sequence of coin flips is an event (that is, has a probability). It's true that many of them do (for instance, "Every flip comes up tails"; "Every even flip comes up heads and every odd flip comes up tails"; "The nth flip comes up tails iff n is divisible by the nth digit of π"; and so on). However, there are in fact very many that don't. To see this, we have only to note that there are uncountably many sequences of coin flips, but only countably many sentences of the agent's language (assuming the agent has a finitary mind). This is where the assumption (in footnote 1) that the events are tied to sentences of the language in some way becomes significant. Whether or not the events are directly tied to sentences of the language however, it at least seems plausible that only countably many of them are distinguishable enough to the agent that they can be considered. This cardinality argument blocks the counter-argument.

In fact, once we notice that the algebra of events is going to be countable (in this framework), we can see that no counterargument of this form will work. Assuming that the language has some amount of metalinguistic capacities, we can consider the event, "Some event of probability 0 occurs". This has just been described in the language, so it is in fact an event as well. In addition, it is coextensive with the disjunction of all events of probability 0. Since there are only countably many events total, there are only countably many of these events. Given countable additivity, this event must itself have probability 0.[3] Since no event with positive probability entails an event with probability 0, each event with positive probability must contain a state that is not in this event. So if we remove all states that are contained within events of probability 0, then we end up with a new probability space in which the only event of probability 0 is empty. So we can salvage a

[3]This gives a kind of impredicativity to the definition, since it must itself be one of the events quantified over in collecting all events of probability 0. However, I think this is an unproblematic impredicativity.

kind of regularity by requiring the algebra of events to be countable.

At this point, the reader may have noticed a slight problem. I have required that the algebra of events be countable, and I have required countable additivity of probabilities for events in the algebra. However, there is no σ-algebra that is countably infinite. It is not hard to prove that if there are countably many distinct events, then there must in fact be infinitely many disjoint events—but then there are uncountably many collections of these, and the countable union of each of those will be a distinct event. Thus, any infinite σ-algebra is uncountable. Since it seems at least rationally permissible (if not required) that there be infinitely many distinct events, this seems to be a problem.

I solve this problem just by conceding that the algebra of events is not in fact a σ-algebra. The reason it is normally required that the algebra be a σ-algebra is that Kolmogorov (and most others) have wanted countable additivity to hold, and countable additivity only makes sense when the union of countably many events is itself an event. In the framework I am considering, there are cases (like the one mentioned above, "Some event of probability 0 occurs") where one event is the union of countably many others. However, because I am requiring that events be describable in the agent's language, there are relatively few countable sets of events whose union is also an event. Instead of *arbitrary* countable intersections and unions of events being events, I will only require that *definable* countable intersections of unions of events be events. If the agent can uniformly describe some countable sequence of events (like above, the sequence H_n of events, "The first n flips all came up heads"), then the agent can easily describe their union (or intersection) as well. "Some event of the form 'The first n flips all came up heads' has occurred." This just relies on some slight metalinguistic capacities for the agent's language. Note that this restriction on countable unions and intersections is exactly what was required to guarantee that not every sequence of flips gets eliminated, when one eliminates all states contained in events of probability 0. Every sequence of flips is the intersection of *some* sequence of events like the following: "Flips $a_0, \ldots, a_i$ all came up heads, and no other flips in the first n did." If every sequence of events of that form had an intersection that was itself an event, then every infinite sequence of flips would be a state whose singleton was an event, and would thus be an event of probability 0. However, only sequences that are uniformly describable (and thus sequences of flips that are describable in the agent's language) end up this way. Countable additivity is only required to hold for countable collections of disjoint events whose union is itself an event.

Another way to think of what is going on here is to give the first-order axiom schemata to define a σ-algebra and note that the Löwenheim-Skolem Theorem proves that there will be countable models of these axioms as well as uncountable ones. (The standard axioms for a σ-algebra are second-order, involving quantification over all countable sets of elements. The first-order axioms will instead include schemata for the nameable countable sets.)[4] There might be some apparent incongruity in requiring countable additivity, but not requiring that every countable collection of events have a union, but I think that is a small price to pay.

The proposal bears an interesting relation to some standard mathematical accounts of randomness—there is a very real sense in which removing the elements of Ω contained in events of probability 0 means that the agent thinks the actual point must be "random." The notion of randomness that is relevant here is that first described in [Martin-Löf, 1966]. Martin-Löf's account of randomness can be described as follows. Consider a real number whose decimal expansion is finitely long, and define the "basic" set corresponding to this number as the set of all real numbers whose decimal expansion begins with this finite sequence. Say that a set of real numbers is "effective" iff it is the union of the sequence of basic sets corresponding to a sequence of finitely long real numbers that can be the sequential outputs of some algorithm. Say that a sequence of effective sets is "constructive" iff the sequence of algorithms generating these sets is itself the sequence of outputs of some further algorithm. A "constructive null cover" of a real number is a constructive sequence of effective sets, each one of which contains that number, such that the limit of the probabilities of the sets is 0. A real number is said to be Martin-Löf random iff it has no constructive null cover. An important result proved in [Martin-Löf, 1966] is the fact that there is actually a universal constructive null cover, so that every non-random real is contained in the same constructive null cover.

For a certain sort of agent, the events of probability 0 correspond exactly to the constructive null covers, so that following the strategy discussed here will mean that Ω consists of exactly the Martin-Löf random reals. The fact that she can consider the event, "Some event of probability 0 has occurred" corresponds to Martin-Löf's theorem on the existence of a universal constructive null cover. For the relevant sort of agent, atomic sentences in her language correspond to (finite unions of) basic sets of real numbers—this is

[4] I owe this point to a conversation with Zach Weber.

very plausible if we accept the motivation that there are limits to human perception, so that any test can only tell the point that a dart hits to some specific finite degree of accuracy, which just tells us some basic set that the dart has hit. If the agent's language is then powerful enough to describe any algorithm, then she will have sentences corresponding to effective sets ("At least one of the following events has occurred," followed by the description of an algorithm for generating the sequence of basic sets) and to the intersections of constructive sequences of effective sets ("All of the following events have occurred," followed by the description of an algorithm for enumerating the effective sets). Thus, she can talk about all constructive null covers, and assign them probability 0, and therefore she must remove all points in these sets. Thus, every remaining point is a Martin-Löf random real.[5]

9.3 Zaman

Asad Zaman has discussed a similar notion in [Zaman, 1987]. Here he uses the fact that in infinitary logic, it is consistent to assume the truth of every conjunct and the falsity of the conjunction, if the conjunction is infinite. This uses the compactness property of standard logics—a contradiction is deducible from a set of premises iff there is some finite subset of the premises from which the contradiction is deducible. The example he uses is of a random variable X uniformly distributed over the interval $[0, 1]$. The probability of the event $X = x$ is 0 for any particular x. For any set S he defines the statement $Q^*(S)$ as the infinitary disjunction of the sentences $X = x$ for all $x \in S$. If we let Ω be the cardinality of the continuum, then we see that it is consistent (though Ω-inconsistent) to assert every sentence of the form $Q^*(S)$ for which S has measure 1, which necessarily means denying every sentence of this form where S has measure 0.

In this setting, he hasn't weakened the collection of events the way I have above—there is still an event for every point, and even for every set of points, not just the describable ones. Now, if an observation is made stating that X does in fact take the value x_0, Zaman says

> After observing $X = x_0$, we appear to face the difficulty of having to change the truth value of $Q(x_0)$ from false to true. We continue to maintain that $Q(x_0)$ is

[5]There are some complications here—the definition of Martin-Löf randomness only uses an algorithmic intersection of algorithmic unions of basic sets, while the agent I have described can layer the processes of algorithmic intersection and union arbitrarily many times. However, I suspect that these processes (unlike general countable intersection and union) can be rearranged so that there is no extra generality here.

> false, but now regard the disjunction $Q^*([x_0 - \epsilon, x_0 + \epsilon])$ as being true for any value of $\epsilon > 0$. This gives us a logically consistent identification of probability 0 with impossibility, and also demonstrates once again the pitfalls of intuitive reasoning about infinities. [Zaman, 1987, p. 158]

This solution thus treats events of the form $X = x$ as infinitary conjunctions of certain events of positive probability. However, although there is no logical contradiction, there is still something very odd about this solution, which relies on the incompleteness of any compact logic. The assertions still cannot all be true under their standard interpretations, even if they could all simultaneously be true.

There is also a further problem, unnoticed by Zaman, that is central here, namely conditional probability. We can use his account to assign conditional truth-values—every set containing an open interval around x_0 is true, and their complements are all false. However, there are other sets whose truth-value is not thus determined, and therefore seem to need a non-extreme conditional probability. For instance, consider the interval $S = [x_0, 1]$—neither it nor its complement contains an interval around x_0. One natural suggestion for its new probability is to take the limit of $P(X \in S | X \in [x_0 - \epsilon, x_0 + \epsilon])$ as ϵ goes to 0, which in this case would give 1/2, which seems reasonable. However, a question arises as to why to take this limit, as opposed to the limit of $P(X \in S | X \in [x_0 - 2\epsilon, x_0 + \epsilon])$, which is also the limit of the conditional probability on a sequence of intervals, and which leads to the value 1/3. This is exactly the problem for the first idea in the previous chapter for conditionalization on events of probability zero.

In this case, we aren't conditionalizing on any particular event (after all, we claim that $X \neq x_0$!), but we are saying what our probabilities should be after learning every event in some infinite family. This is an instance of the distinction I drew in Chapter 3 between updating and conditional probability. Because this infinite family can be learned in many different ways, the result is indeterminate.

To avoid this sort of problem, my defender of the ratio analysis will have to go farther. The agent rules out all states that are contained in some event of probability 0–thus, she can't conditionalize on any such event. Instead, she will only ever be able to conditionalize on events of positive probability. This prevents her from ever having to worry about dividing by zero, but it means that the events she conditionalizes on must always have positive probability. In particular, this means that she must only ever learn such propositions, but it has further consequences for other applications of conditional probability.

9.4 Empiricism

Although this approach has some intriguing features, I think it is too committed to a particular picture of human perceptual capacities. Since the agent is committed to never conditionalizing on events of probability 0, she is prevented from Jeffrey conditionalizing on a continuous partition, the way the account I favor in the previous chapters allows. Instead, her update procedure must either be traditional conditionalization (on some event of non-zero probability) or Jeffrey conditionalizing on a partition of events of non-zero probability. Apart from avoiding the issues of conditionalizing on events of probability 0, the only motivation I can see for this idea is an appeal to the finite limits of human discrimination. It's often suggested that in the sorts of situations that give rise to the standard examples of conditioning on events of probability 0, (which often involve learning one coordinate of the location of some point given by multiple coordinates) the limits of human discrimination mean that any observation must be an observation of some *interval* containing the actual value of the coordinate, rather than the specific value itself.

This response must apply to all uses of conditional probabilities, and not just to updates. Thus, my defender of the ratio account of conditional probability must say something similar about the conditional probabilities in confirmation theory and decision theory as well. For decision theory this may be very plausible—perhaps agents really only ever can distinguish finitely many possible actions, and thus every action they contemplate has positive probability, so the ratio account works. But for confirmation, the problem is trickier. Whatever experiment the agent contemplates performing, the relevant possible outcomes must all have positive probability for her. Therefore, either there must be some sort of finite partition of the possible outcomes into indistinguishability classes, or else the outcomes must not form a partition.

The former suggestion seems very strange. If the agent is contemplating an observation of the position of some arrow on a spinner, it seems strange to say that these outcomes can be partitioned into events of positive probability in any reasonable way. There are clearly no particular points at which the spinner might be, such that being any farther to the left would cause an apparent "jump" to the next class of indistinguishable points. Allowing the possible outcomes not to form a partition might work better, though it's hard to say exactly what it would mean to observe the point falling in one interval, rather than a narrower or wider interval. There's still a sort of arbitrariness in deciding exactly what the limits to our perception are. The more natural approach to dealing with this sort of limit to our perceptual abilities is by means of Jeffrey conditionalization, where the original distribution is uniform around the spinner, and the new distribution is concentrated around some point. But this seems to require a partition into miniscule events. Of course, this problem is even worse if the relevant conditional probability is a likelihood rather than a posterior—this sort of agent could not conditionalize on particular settings of a parameter in a theory, but must conditionalize on some interval.

9.5 Conclusions

Instead, I think a more flexible solution is that proposed in the previous chapter, of letting conditional probability be defined by a function $P(A|B, \mathcal{E})$, defined on a state space Ω and algebra $\mathcal{A}$. But there is still room for more work to determine exactly which partitions $\mathcal{E}$ are the relevant ones to use, and how in general to compute the values of these conditional probabilities (if in fact there is any unique constraint given to them by the values of unconditional probabilities). I don't claim to have given a complete theory of conditional and unconditional degree of belief and their relation—I have just investigated the foundational connections between these two notions and shown that they must be related in some ways that are unexpected.

I don't claim that the principles I support relating conditional and unconditional probability form the best set of axioms for these notions—I just claim that these relations are rationally required. [Goosens, 1979] and others point out that a proper axiomatization is an important goal in understanding probability—I agree, but suspect that justifying one set of axioms over another is a very subtle prospect that requires having established the proper theory first. (Consider the discussion about whether the T-schema is best considered as a set of axioms for truth, or merely a set of theorems that any adequate theory of truth must prove—it was necessary to first establish the importance of this schema.) The debate about whether conditional or unconditional probability is the more fundamental notion (as discussed in [Hájek, 2003]) will eventually require a settling of this question of axiomatization, but I think providing certain aspects of the theory can at least give strong hints as to what this relation might end up being.

Future projects developing from this one would involve figuring out whether further constraints must hold for conditional and unconditional probability. I have merely defended enough constraints to run my central arguments—there may be more that are required. (For instance, some more general version of a Principle of Indifference may hold, as suggested by [Jaynes, 2003], or something like the Principal Principle introduced in [Lewis, 1980].) I also don't give a full account of how rational agents ought to update their doxastic state. I suggest that a version of Jeffrey conditionalization is often the right way to do it, but I have not given any account of what makes one Jeffrey update the right one rather than another. I also don't deal with the problem of forgetting, or learning essentially indexical information. (These problems are addressed in [Titelbaum, forthcoming], though it remains to be seen how to make his account there compatible with my general account of conditional probability.)

There may also be further objections raised to some of the arguments I make. Some of my claims may end up being relatively independent of one another—I have defended all of them because they seem essential to the particular method I use to justify my main claims. However, it's possible that other arguments for the same conclusions will not rely on all the claims that I make.

Appendix A

A proof of the integral equation from conglomerability

Let $\mathcal{E}$ be some partition of a probability space into disjoint events E_α. Let $f_A(w) = P(A|E_\alpha)$, where E_α is the unique element of the partition containing the point w. I want to show that for any B that is the union of some collection of E_α, $P(A\&B) = \int_B f_A(w)dw$.

Note that the integral used here is the Lebesgue integral, which is defined as the supremum of the sums $\sum x_i P(h(w) = x_i)$ over measurable functions h that are 0 outside B, everywhere bounded above by f, and take on only finitely many distinct values x_i. I will abuse notation and use the integral symbol for the sum in dealing with functions that only take finitely many values. That is, for such a function h, I will define $\int h(w)dw = \sum x_i P(h(w) = x_i)$.

It is clear that if h' is some function 0 outside B, everywhere bounded *below* by f, taking on only finitely many distinct values x_i', then $\int_B f_A(w)dw \leq \int h'(w)dw$, because for any h considered in calculating $\int f_A(w)dw$, we know that $h(w)$ is always at $h'(w)$, so $\int h(w)dw \leq \int h'(w)dw$.

Thus, if for every n I can find h_n and h_n' that are 0 outside B and such that everywhere in B, $h_n \leq f \leq h_n'$ and $\int h_n(w)dw \leq P(A\&B) \leq \int h_n'(w)dw$ and $\int h_n'(w)dw - \int h_n(w)dw \leq 1/n$, then I will have proven the integral equation. This is because the integral of h_n' is an upper bound for the integral of f and the integral of h_n is a lower bound, and since n is arbitrary, they can both be made arbitrarily close to $P(A\&B)$.

So now let $h_n(w)$ and $h_n'(w)$ both be 0 for $w \notin B$ and let $h_n(w) = \max\{\frac{k}{n} : \frac{k}{n} \leq$

$f_A(w)\}$ and $h'_n(w) = h_n(w) + 1/n$ for $w \in B$. Because $f_A(w)$ was defined in terms of which element E_α of the partition contained w, we see that f is constant on the E_α, so h_n and h'_n are too. Thus, the set B_k where $h_n(w) = \frac{k}{n}$ (which is also where $h'_n(w) = \frac{k+1}{n}$) is a union of some of the E_α. In particular, it is the union of the E_α where $\frac{k}{n} \leq P(A|E_\alpha) < \frac{k+1}{n}$.

Thus, by the strong version of conglomerability, we see that $\frac{k}{n} \leq P(A|B_k) \leq \frac{k+1}{n}$. Multiplying through by $P(B_k)$, we get $\frac{k}{n}P(B_k) \leq P(A\&B_k) \leq \frac{k+1}{n}P(B_k)$. Summing over all k from 0 to n (these are the only relevant values, because f was bounded between 0 and 1), we get $\int h_n(w)dw \leq P(A\&B) \leq \int h'_n(w)dw = \int h_n(w)dw + P(B)/n$. But this is just what we wanted earlier, since $P(B) \leq 1$.

Thus, the argument goes through as desired.

Appendix B

A construction that avoids relativization in the Borel paradox

In this appendix I will show how Kolmogorov's extended analysis of conditional probabilities can be used to define every conditional probability in the Borel paradox absolutely, rather than just relative to a particular axis. However, this non-relativized conditional probability function will be highly non-unique, and specifying any particular such function will require a well-ordering of the reals. In addition, this proof requires controversial set-theoretic principles beyond just the Axiom of Choice. Because of these problems, I will endorse the relativized function of the main body of the text, rather than this absolute function, which may not even exist for many spaces.

Recall that if we consider a particular axis, then relative to the partition of the sphere into the great circles through this particular axis, the conditional probability of any set given any great circle is $(\cos\theta_0 - \cos\theta_1)/4$ if it intersects the great circle in the interval from θ_0 to θ_1 radians away. It isn't necessary that every great circle have this property for all its conditional probabilities, just that for any axis, the set of great circles that do should together form a set of measure 1 (by the result of the previous appendix). Since these values depend greatly on the point from which the angles are measured, each great circle can only give all the "correct" values for at most one axis it goes through. Thus, in order to define these conditional probabilities absolutely, I will associate each great circle with a single axis it goes through in such a way that the set of great circles associated with any particular axis covers a region of measure 1 on the surface of the sphere.

To do this, I will assume the Continuum Hypothesis (that any set of real numbers with cardinality strictly less than that of the set of all real numbers is in fact countable)[1]. There are just as many possible axes for a sphere as there are real numbers. So using the Axiom of Choice, let us well-order them in the shortest order-type possible, so that (by the Continuum Hypothesis), every axis has only countably many predecessors in the ordering. Then associate each great circle with the axis on it that comes earliest in the well-ordering. Thus, each great circle will be associated with a unique axis it goes through, so we can define the probability of any set conditional on this great circle to be the value it should have relative to this particular axis.

Since each circle is associated with the earliest axis on it in the well-ordering, the only way a particular great circle through an axis A can be unassociated with A is if it contains an axis that comes earlier than A in the well-ordering. But as mentioned above, there are only countably many axes earlier than A, and any pair of axes have exactly one great circle going through both of them, so there are at most countably many great circles through A that aren't associated with A. By countable additivity, the great circles not associated with A have measure 0, so the ones associated with axis A have measure 1. Since this is true for any axis A, the association of great circles with axes has the property required above, QED.

This result is highly counterintuitive. It says that although any given great circle gives the "wrong" probabilities for almost every axis on it, we can make sure that for any axis, almost every great circle through it gives the "right" probabilities. I would argue that this highlights some counterintuitive aspects of the Continuum Hypothesis when combined with the Axiom of Choice.

Using this association of circles with axes, we can define all the conditional probabilities discussed above absolutely, rather than relatively, in a way that preserves all the integrals needed. However, it is not clear that it is possible to extend this definition to probabilities conditional on other sets of measure zero (say, lines of latitude rather than longitude). It is also unclear whether these probabilities will satisfy the appropriate inte-

[1] This proof will in fact go through using Martin's Axiom, which is weaker than the Continuum Hypothesis. Both are known to be consistent with, but independent from, standard ZFC set theory. Using Martin's Axiom, every cardinal smaller than the number of real numbers behaves like $\aleph_0$, and in particular, the union of κ many sets of measure 0 is itself a set of measure 0 under standard Lebesgue measure, assuming that $\kappa < 2^{\aleph_0}$. For more information on these axioms, see [Kunen, 1980, p. 51]: "Unlike the basic axioms of ZFC, MA does not pretend to be an 'intuitively evident' principle, and in fact at first sight it seems strange and ill-motivated."

gral equations relative to other partitions of the space not considered here.

More importantly, the conditional probabilities so defined are highly non-uniform, and depend on the well-ordering of the set of axes. Such a well-ordering can only be given in a highly non-constructive way, using the Axiom of Choice, and there are many more such well-orderings than there are real numbers. Each well-ordering gives a different set of conditional probabilities, so it is particularly hard to justify any set of these as the "correct" set of conditional probabilities for this example. But, as I argued in section 4.2.1, since there is no fact about the agent's mind that could distinguish these well-orderings, there can't be any reason for one of these to be the "correct" one.

Index

Bibliography

[Arntzenius et al., 2004] Arntzenius, F., Elga, A., and Hawthorne, J. (2004). Bayesianism, infinite decisions, and binding. *Mind*, 113(450):251–283.

[Barnett, 2007] Barnett, D. Ramsey + Moore $\neq$ God. Unpublished ms.

[Bartha, 2004] Bartha, P. (2004). Countable additivity and the de Finetti lottery. *British Journal for the Philosophy of Science*, 55:301–321.

[Bartha and Johns, 2001] Bartha, P. and Johns, R. (2001). Probability and symmetry. *Philosophy of Science*, 68:S109–S122.

[Bennett, 2003] Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford.

[Bertrand, 1889] Bertrand, J. (1889). *Calcul des probabilités*. Gauthier-Villars.

[Carnap, 1945] Carnap, R. (1945). Two concepts of probability. *Philosophy and Phenomenological Research*, 5(4):513–532.

[Carnap, 1950] Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press.

[Chalmers, 2002] Chalmers, D. (2002). The St. Petersburg two-envelope paradox. *Analysis*, 62(2):155–157.

[Chalmers and Hájek, 2007] Chalmers, D. and Hájek, A. (2007). Ramsey + Moore = God. *Analysis*, 67(72):170–172.

[Christensen, 1996] Christensen, D. (1996). Dutch book arguments de-pragmatized: Epistemic consistency for partial believers. *The Journal of Philosophy*, 93(9):450–479.

[Christensen, 1999] Christensen, D. (1999). Measuring confirmation. *The Journal of Philosophy*, 96(9):437–461.

[Christensen, 2001] Christensen, D. (2001). Preference-based arguments for probabilism. *Philosophy of Science*, 68(3):356–376.

[Colyvan, 2004] Colyvan, M. (2004). The philosophical significance of Cox's theorem. *International Journal of Approximate Reasoning*, 37(1):71–85.

[Cox, 1946] Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13.

[Deutsch, 1999] Deutsch, D. (1999). Quantum theory of probability and decisions. *Proceedings of the Royal Society A*, 455(1988):3129–3137.

[Dubins, 1975] Dubins, L. (1975). Finitely additive conditional probabilities, conglomerability, and disintegrations. *The Annals of Probability*, 3(1):89–99.

[Earman, 1992] Earman, J. (1992). *Bayes or Bust?* The MIT Press.

[Eells and Fitelson, 2000] Eells, E. and Fitelson, B. (2000). Measuring confirmation and evidence. *The Journal of Philosophy*, 97(12):663–672.

[Eells and Skyrms, 1994] Eells, E. and Skyrms, B., editors (1994). *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge University Press.

[Elga, 2000] Elga, A. (2000). Self-locating belief and the Sleeping Beauty problem. *Analysis*, 60(2):143–147.

[Eriksson and Hájek, 2007] Eriksson, L. and Hájek, A. (2007). What are degrees of belief? *Studia Logica*, 86:185–215.

[Fara, 2006] Fara, M. (2006). Dispositions. *Stanford Encyclopedia of Philosophy*.

[Feller, 1968] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. John Wiley & Sons.

[Field, 1977] Field, H. (1977). Logic, meaning, and conceptual role. *The Journal of Philosophy*, 74(7):379–409.

[Fitelson, 1999] Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, 66(3):S362–S378.

[Gaifman, 2004] Gaifman, H. (2004). Reasoning with limited resources and assigning probabilities to arithmetical statements. *Synthese*, 140:97–119.

[Garber, 1983] Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. In Earman, J., editor, *Testing Scientific Theories*, volume 10. Minnesota Studies in the Philosophy of Science.

[Gillies, 2000] Gillies, D. (2000). Varieties of propensity. *British Journal for the Philosophy of Science*, 51:807–835.

[Glymour, 1980] Glymour, C. (1980). *Theory and Evidence*. Princeton University Press.

[Goel and Zellner, 1986] Goel, P. K. and Zellner, A., editors (1986). *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*. North-Holland.

[Goldblatt, 1989] Goldblatt, R. (1989). *Lectures on the Hyperreals: an introduction to nonstandard analysis*. Graduate Texts in Mathematics. Springer-Verlag.

[Goosens, 1979] Goosens, W. (1979). Alternative axiomatizations of elementary probability theory. *Notre Dame Journal of Formal Logic*, 20(1):227–239.

[Greaves and Wallace, 2006] Greaves, H. and Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115(459):607–632. MS.

[Hailperin, 1984] Hailperin, T. (1984). Probability logic. *Notre Dame Journal of Formal Logic*, 25(3):198–212.

[Hailperin, 1997] Hailperin, T. (1997). Ontologically neutral logic. *History and Philosophy of Logic*, 18:185–200.

[Hailperin, 2000] Hailperin, T. (2000). Probability semantics for quantifier logic. *Journal of Philosophical Logic*, 29:207–239.

[Hailperin, 2007] Hailperin, T. (2007). Quantifier probability logic and the confirmation paradox. *History and Philosophy of Logic*, 28:83–100.

[Hájek, 2003] Hájek, A. (2003). What conditional probability could not be. *Synthese*, 137:273–323.

[Hájek, 2007a] Hájek, A. (2007a). Dutch book arguments. Unpublished ms.

[Hájek, 2007b] Hájek, A. (2007b). Interpretations of probability. *Stanford Encyclopedia of Philosophy*.

[Halmos, 1974] Halmos, P. (1974). *Lectures on Boolean Algebras*. Springer-Verlag.

[Halpern, 2004] Halpern, J. (2004). Lexicographic probability, conditional probability, and nonstandard probability. ArXiV, July 10, 2004.

[Hill and Lane, 1986] Hill, B. M. and Lane, D. (1986). Conglomerability and countable additivity. In *[Goel and Zellner, 1986]*.

[Howson, 2008] Howson, C. (2008). De Finetti, countable additivity, consistency and coherence. *The British Journal for the Philosophy of Science*, 59(1):1–23.

[Humphreys, 1985] Humphreys, P. (1985). Why propensities cannot be probabilities. *The Philosophical Review*, 94(4):557–570.

[Humphreys, 2004] Humphreys, P. (2004). Some considerations on conditional chances. *British Journal for the Philosophy of Science*, 55:667–680.

[Jaynes, 1973] Jaynes, E. T. (1973). The well-posed problem. *Foundations of Physics*, 3(4):477–492.

[Jaynes, 2003] Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

[Jech, 2003] Jech, T. (2003). *Set Theory: The Third Millenium Edition, Revised and Expanded*. Springer-Verlag.

[Joyce, 1998] Joyce, J. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4):575–603.

[Joyce, 1999] Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

[Kadane et al., 1986] Kadane, J. B., Schervish, M. J., and Seidenfeld, T. (1986). Statistical implications of finitely additive probability. In *[Goel and Zellner, 1986]*.

[Kolmogorov, 1950] Kolmogorov, A. N. (1950). *Foundations of the Theory of Probability*. Chelsea.

[Kunen, 1980] Kunen, K. (1980). *Set Theory: an introduction to independence proofs*. North-Holland.

[Kyburg, 1978] Kyburg, H. (1978). Subjective probability: Criticisms, reflections, and problems. *Journal of Philosophical Logic*, 7:157–180.

[Langendoen and Postal, 1984] Langendoen, D. T. and Postal, P. (1984). *The Vastness of Natural Languages*. Basil Blackwell.

[Lange, 2006] Lange, M. (2006). Do chances receive equal treatment under the laws? or: Must chances be probabilities? *British Journal for the Philosophy of Science*, 57:383–403.

[Lewis, 1970] Lewis, D. (1970). How to define theoretical terms. *The Journal of Philosophy*, 67(13):427–446.

[Lewis, 1980] Lewis, D. (1980). A subjectivist's guide to objective chance. In Jeffrey, R., editor, *Studies in Inductive Logic and Probability*, volume II. University of California Press.

[Maher, 1996] Maher, P. (1996). Subjective and objective confirmation. *Philosophy of Science*, 63(2):149–174.

[Maher, 2006a] Maher, P. (2006a). The concept of inductive probability. *Erkenntnis*, 65:185–206.

[Maher, 2006b] Maher, P. (2006b). A conception of inductive logic. *Philosophy of Science*, 73:518–523.

[Martin-Löf, 1966] Martin-Löf, P. (1966). The definition of random sequences. *Information and Control*, 9:602–619.

[McCurdy, 1996] McCurdy, C. (1996). Humphreys's paradox and the interpretation of inverse conditional probabilities. *Synthese*, 108:105–125.

[McGee, 1994] McGee, V. (1994). Learning the impossible. In Eells, E. and Skyrms, B., editors, *Probability and Conditionals*. Cambridge University Press.

[McGee, 1999] McGee, V. (1999). An airtight Dutch book. *Analysis*, 59(4):257–265.

[Meek and Glymour, 1994] Meek, C. and Glymour, C. (1994). Conditioning and intervening. *British Journal for the Philosophy of Science*, 45:1001–1021.

[Popper, 1959a] Popper, K. (1959a). *The Logic of Scientific Discovery*, chapter iv*, pages 326–348. Harper & Row.

[Popper, 1959b] Popper, K. (1959b). The propensity interpretation of probability. *The British Journal for the Philosophy of Science*, pages 25–42.

[Ramsey, 1926] Ramsey, F. P. (1926). Truth and probability. In Braithwaite, R. B., editor, *The Foundations of Mathematics and other Logical Essays (1931)*. Harcourt, Brace and Company.

[Rényi, 1970] Rényi, A. (1970). *Foundations of Probability*. Holden-Day.

[Robinson, 1996] Robinson, A. (1996). *Non-standard analysis*. Princeton University Press.

[Roeper and LeBlanc, 1999] Roeper, P. and LeBlanc, H. (1999). *Probability Theory and Probability Logic*. University of Toronto.

[Salmon, 1967] Salmon, W. (1967). *The Foundations of Scientific Inference*. University of Pittsburgh Press.

[Schaffer, 2007a] Schaffer, J. (2007a). Deterministic chance? *British Journal for the Philosophy of Science*, 58:113–140.

[Schaffer, 2007b] Schaffer, J. (2007b). Knowing the answer. *Philosophy and Phenomenological Research*, 75(2):383–403.

[Schervish et al., 1984] Schervish, M. J., Seidenfeld, T., and Kadane, J. B. (1984). The extent of non-conglomerability of finitely additive probabilities. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 66:205–226.

[Seidenfeld et al., 2001] Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (2001). Improper regular conditional distributions. *The Annals of Probability*, 29(4):1612–1624.

[Shafer and Vovk, 2005] Shafer, G. and Vovk, V. (2005). The origins and legacy of Kolmogorov's *Grundbegriffe*.

[Strevens, 1998] Strevens, M. (1998). Inferring probabilities from symmetries. *Noûs*, 32(2):231–246.

[Tenenbaum and Griffiths, 2001] Tenenbaum, J. and Griffiths, T. (2001). Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences*, 24:629–640.

[Titelbaum, forthcoming] Titelbaum, M. The relevance of self-locating beliefs.

[Tversky and Kahneman, 1983] Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315.

[van Fraassen, 1977] van Fraassen, B. (1977). Relative frequencies. *Synthese*, pages 133–166.

[van Fraassen, 1980] van Fraassen, B. (1980). Critical notice of Brian Ellis, *Rational Belief Systems*. *Canadian Journal of Philosophy*, 10(3):497–511.

[van Fraassen, 1984] van Fraassen, B. (1984). Belief and the will. *The Journal of Philosophy*, 81(5):235–256.

[van Fraassen, 1995] van Fraassen, B. (1995). Belief and the problem of Ulysses and the Sirens. *Philosophical Studies*, 77:7–37.

[Weatherson, 2003] Weatherson, B. (2003). From classical to intuitionistic probability. *Notre Dame Journal of Formal Logic*, 44(2):111–123.

[Williamson, 1999] Williamson, J. (1999). Countable additivity and subjective probability. *The British Journal for the Philosophy of Science*, 50:401–416.

[Williamson, 2002] Williamson, T. (2002). *Knowledge and its Limits*. Oxford.

[Williamson, 2007] Williamson, T. (2007). How probable is an infinite sequence of heads? *Analysis*, 67(3):173–180.

[Zaman, 1987] Zaman, A. (1987). On the impossibility of events of zero probability. *Theory and Decision*, 23:157–159.