Scientific Uncertainty and Decision Making

Seamus Bradley

A thesis submitted to the Department of Philosophy, Logic and Scientific Method of the London School of Economics for the degree of Doctor of Philosophy, London, September 2012

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Part of section 3.1 has been published as Bradley (2012). A version of chapter 6 has been published as Bradley (2011).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 95 497 words.

Dedication

Thanks to all the people at conferences with whom I have discussed my ideas. Thanks especially to Chris Clarke, Jonny Blamey, Alastair Wilson, David Etlin, Richard Pettigrew, Wendy Parker, Robbie Williams and Alan Hájek. Thanks to the students and staff at the LSE who have been incredibly helpful. Thanks to Dean Peters, Foad Dizadji-Bahmani, Conrad Heilmann, Ittay Nissan. Thanks to Katie Steele, Lenny Smith and Dave Stainforth for comments on parts of drafts. Thanks to my supervisors Richard Bradley and Roman Frigg for support and encouragement.

Thanks also to my family for their support. Above all, thank you Megan, for keeping me (almost) sane.

Abstract

It is important to have an adequate model of uncertainty, since decisions must be made before the uncertainty can be resolved. For instance, flood defenses must be designed before we know the future distribution of flood events. It is standardly assumed that probability theory offers the best model of uncertain information. I think there are reasons to be sceptical of this claim.

I criticise some arguments for the claim that probability theory is the only adequate model of uncertainty. In particular I critique Dutch book arguments, representation theorems, and accuracy based arguments.

Then I put forward my preferred model: imprecise probabilities. These are sets of probability measures. I offer several motivations for this model of uncertain belief, and suggest a number of interpretations of the framework. I also defend the model against some criticisms, including the so-called problem of dilation.

I apply this framework to decision problems in the abstract. I discuss some decision rules from the literature including Levi's E-admissibility and the more permissive rule favoured by Walley, among others. I then point towards some applications to climate decisions. My conclusions are largely negative: decision making under such severe uncertainty is inevitably difficult.

I finish with a case study of scientific uncertainty. Climate modellers attempt to offer probabilistic forecasts of future climate change. There is reason to be sceptical that the model probabilities offered really do reflect the chances of future climate change, at least at regional scales and long lead times. Indeed, scientific uncertainty is multi-dimensional, and difficult to quantify. I argue that probability theory is not an adequate representation of the kinds of severe uncertainty that arise in some areas in science. I claim that this requires that we look for a better framework for modelling uncertainty.

Contents

List of Figures

1. Introduction

> But whatever Socrates may say, it remains the case, as anyone can see, that people who stick to philosophy become strange monsters, not to say utter rogues; even the best of them are made useless by philosophy.
>
> *(Bertrand Russell)*

Uncertainty is ubiquitous. It is important to understand what uncertainty is, and where it comes from. It is also important to know how to reason about uncertainty, and particularly how to make decisions when faced with uncertainty. The standard set of formal methods for dealing with uncertainty is built around the mathematical theory of probability. This has been a remarkably successful theory, with a huge number of applications. I will argue that there are circumstances where probability theory is too restrictive a representation of uncertainty. There are alternatives to probability theory. I will argue that in cases of severe uncertainty, they are better suited to represent uncertainty than probability theory is. More generally, I want to argue that we shouldn't expect rationality to determine the right course of action in every circumstance. In cases of severe uncertainty, it might be that all that rationality can do is rule out some options as impermissible.

This is an important project because one area where severe uncertainty is prevalent is in the field of climate science. Decisions of global importance need to be made that rely on scientific evidence from climate science. Thus, it is important to properly represent the uncertainty present in order to facilitate good decisions. The standard precise probabilistic methodology might suggest an unwarranted level of certainty. This certainty could encourage decision makers to attempt to optimise their responses. But such optimisation with possibly faulty probabilities does not give any guarantee of success.

Before we can start a study of the alternatives to the orthodox probabilistic theory, it will be instructive to study that theory in detail. I outline the formal preliminaries and introduce the central concepts of the dissertation in chapter 2. There will be some work to do here in justifying particular choices I make about how to set up the framework. Many arguments have been put forward that purport to show that probability theory is the best model of uncertain reasoning and decision making. Call the view that probability theory is the best model of uncertain reasoning "Probabilism". So a prerequisite for this project is that we show how and why these arguments fail or are not appropriate in the current context. This is the aim of chapter 3.

I do not argue that probability theory is never an appropriate model of uncertainty. Often it is a very good one, and I hope to be conservative in the ways I go beyond the standard theory. I discuss a variety of methods for going beyond orthodox probability theory in chapter 4. Various arguments for my preferred framework will be offered, and two problems with learning in this *imprecise probability* framework will be discussed. My favoured framework will be used in chapter 5 to discuss decision making under severe uncertainty. The conclusions are somewhat negative. Decision making under severe uncertainty is difficult, perhaps unavoidably so. It is, however, important to recognise that difficulty, and to not be misled into making poor decisions by over-precise estimates of uncertainty. More generally I think it is asking too much of rationality to always determine a fully precise probability function that represents your uncertainty, whatever the state of your evidence. Furthermore, I think that in situations of severe uncertainty it is not a good idea to attempt to look for optimal courses of action. Optimising on weak evidence like this isn't necessarily a good idea.

I will offer a case study of scientific uncertainty in chapter 6. The claim is that given the multitude of sources of uncertainty and error, perhaps we would be better off using a more permissive formal framework for reasoning about the uncertain scientific propositions. Finally, chapter 7 returns to the issue of decision making and climate change, and the particular problems that we encounter there.

1.1. Characterisations of uncertainty

I should mention something about my use of the word "uncertainty". I am using this as a catch-all term for all the ways one might fail to be certain about things. Economists and decision theorists sometimes make a distinction between "risk" where the probabilities are known, and "uncertainty" where they are not. This distinction is traced back to the work of Frank Knight (Knight 1921). Unfortunately, physicists have the opposite view: "uncertainty" connotes a situation where you know the probabilities and "ambiguity" is the term for a situation where the probabilities are unknown. I use uncertainty to cover both of these possibilities. The terminology is not as stable as this suggests, but to the extent that it is stable, it has stabilised in incompatible ways in different disciplines.

There have been several attempts to classify kinds of uncertainty. For example Walker et al. (2003) offer a framework for understanding uncertainty when using evidence from models to make decisions. They identify three distinct "axes" of uncertainty of relevance for "model-based decision support". These are:

Location Where exactly the error enters into the process

Level How much the uncertainty affects the model-relevant predictions

Nature Whether the uncertainty is due to a deficiency in the modelling, or due to unavoidable natural variation

The National Institute of Standards and Technology's *Guidelines for Evaluating and Expressing Uncertainty* also note the distinction between *Random* and *Systematic* errors (Taylor and Kuyatt 1997). This is a distinction that falls under the Walker et al. category of *nature of uncertainty*. The NIST document notes that this distinction often matches up with a difference in how the uncertainty can be reported: random errors are often evaluated statistically, systematic errors are not. They also note that the categories are somewhat dependent on perspective:

> The nature of an uncertainty component is conditioned by the use made of the corresponding quantity, that is, on how that quantity appears in the mathematical model that describes the measurement process. When the corresponding quantity is used in a different way, a "random" component may become "systematic" and vice versa.
>
> Taylor and Kuyatt (1997, p. 2)

On a similar note, Lo and Mueller (2010) distinguish a number of types of uncertainty, based on what techniques can be used to deal with them and what amount of data you need to overcome them. The view of physics they present is rather naïve, but their main aim is to draw a distinction between the "easy" cases in physics and the "hard" cases in economics and finance. They take uncertainty in physics to be a lot more clear-cut than it often is; they are keen to stress that finance and economics are much less certain – riddled with more severe uncertainties –

than physics. In chapter 6 I will argue that many of the kinds of uncertainty that Lo and Mueller think are worrying in finance and economics are also troubling in areas like climate science. There may be differences in the severity of the uncertainties encountered in physics and in finance, but it is a difference of degree, not a difference in kind. They distinguish what is often called "risk" (known probabilities) from uncertainty. They then go on to distinguish three grades of uncertainty. The first they call *reducible uncertainty* which is uncertainty where the state space is known, and it is reasonable to assume a fixed stable background. Classical statistics uses observations to reduce the uncertainty to risk. Below this we have *partially reducible uncertainty* where the above conditions no longer hold straightforwardly. There might be time-varying parameters, model uncertainty, non-linearities and so on. In short, there may be aspects of the target system that render invalid the standard statistical results necessary to warrant statistical inference (for example the Law of Large Numbers or the Central Limit Theorem). At this level, more sophisticated modelling techniques will be necessary, perhaps only a coarse-grained model of the phenomena is possible. There are however enough regularities in the target system that progress can be made, however piecemeal and tentative. Below this there is the level of irreducible uncertainty. Irreducible uncertainty is our epistemic situation with respect to those things that are not governed by any regularities or statistical laws at all. No amount of data, no sophistication of technique is enough to accommodate things that fall in this category. As Lo and Mueller point out,

> irreducible uncertainty seems more likely to be the exception rather than the rule. After all, what kinds of phenomena are completely impervious to quantitative analysis?. . . The usefulness of the concept is precisely in its extremity. By defining a category of uncertainty that cannot be reduced to quantifiable risk – essentially an admission of intellectual defeat – we force ourselves to stretch our imaginations to their absolute limits before relegating any phenomenon to this level.
>
> Lo and Mueller (2010, p. 16)

These categories are not wholly distinct, nor are they unrevisable. Solar eclipses would have been irreducibly uncertain for primitive tribes, but they can now be predicted with such accuracy that no uncertainty remains. Assigning something to some category or another will be relative to a state of information: more evidence, better theories or better model techniques might change what level something sits at. This is pushing in the same direction as the NIST *Guidelines* passage quoted earlier. For the statistics to help at the level of reducible uncertainty, for instance, we need to know things like that there is a stable background. But on reflection, we might take Hume's worry about induction seriously and conclude that we can never be certain of that sort of thing. This undermines pretty much every reduction of uncertainty to risk. What we take ourselves to know and to believe depends in part on what we take for granted. Consider betting on the roll of some dice. If we take for granted that the dice are fair, then this is a case of risk. If however we don't take this for granted, we are in a case of reducible uncertainty. Nothing about the dice has changed. The difference in epistemic situation is due to what is taken for granted. What decides what can be taken for granted in a given situation is something extra-rational: we must appeal to some sort of Duhemian *Scientific Good Sense* (Duhem 1998) and leave it at that.

Kandlikar, Risbey, and Dessai (2005) discuss how to represent different kinds of uncertainty. We can understand this as a cashing out of the *level of uncertainty* category from Walker et al. (2003). The best case scenario is that you know enough to produce a probability density function (PDF) that encapsulates all there is to know about the uncertain process at issue. Failing this, perhaps you can only put upper and lower bounds on the possible values of the variables of interest. If even this isn't possible, then you might still be able to assess the order of magnitude of the expected effect. The lowest useful level of information is just to say whether the change in the variable will be positive or negative (the *sign* of the effect): whether the trend is upwards or downwards. If none of these factors can be usefully drawn from the evidence, then you are effectively in a state of complete ignorance.

Regan, Colyvan, and Burgman (2002) offer a typology of uncertainty for ecologists. They define two broad kinds of uncertainty: *Epistemic* and *Linguistic*. Epistemic uncertainty is due to unknown factors in the world; linguistic uncertainty is uncertainty we have in virtue of the inexact way language matches up with nature. I am particularly interested in the epistemic uncertainties, and my characterisation of epistemic uncertainties is not all that different from theirs.

Understanding the nature and diversity of scientific uncertainty is important for policy making. Oversimplifying scientific evidence, or downplaying the uncertainties involved is to the detriment of model based decision making (Covey 2000; Stirling 2010). This is an important theme for the thesis: an adequate understanding of uncertainty must be communicated to – and used by – the decision makers.

It is often assumed that probability theory offers the right representation of uncertain information. The aim of this thesis is to explore this idea. Given the above diversity of kinds of uncertainty and severity of uncertainty, perhaps there is room for other formal frameworks.

1.2. General lessons

I would like to briefly set out the aims of the current dissertation.

The first important lesson I would like to draw out in this dissertation is that uncertainty comes in many forms. Chapter 6 makes this point explicitly for the case of climate science uncertainty. But the multifariousness of uncertainty is an important motivation for exploring extensions of probability theory (chapter 4). And indeed, the multidimensional nature of uncertainty is one important reason to be sceptical of those arguments that purport to show that probability theory is the correct formal model of uncertainty (chapter 3).

A consequence of the myriad forms of uncertainty is that representing uncertainty is important for good decision making. Furthermore, we need this representation of uncertainty to be part of what decision makers use to make their decisions. In section 4.3.4 I give an example of how an inadequate representation of the state of your evidence can lead to an intuitively bad decision. Related to this point is the idea that we aren't always in a position to choose optimal acts. In cases of severe uncertainty, we can't know what is "really" optimal, and what appears optimal given our incomplete evidence needn't be a good option. Adequate representation of severe uncertainty blocks these "premature optimisations". This is an important point to bear in mind when it comes to making decisions based on (uncertain) climate evidence (chapter 7).

My most general claim, the one that perhaps best brings together the disparate chapters of this work is this: we need a subtler understanding of what rationality is, and what we can expect of it. Rationality is a theme that runs through this dissertation: what is it rational to believe?; what is it rational to choose? Rationality is often taken to be part of what determines the answer to these questions. I think in the case of severe uncertainty, we shouldn't expect there to be determinate answers to these questions. So what rationality is doing is simply ruling out some of the options as *irrational*. This shift in understanding of rationality allows us to focus on what conditions are required for rationality to do more. Imprecise probabilism is a representation of uncertainty that is only definite to the extent that the evidence sanctions definite degree of belief (chapter 4). This is a formal model of uncertainty that is true to the idea that rationality needn't always give determinate answers. Once we commit to moving beyond the standard probability theory framework, we need to look at what decision making looks like (chapter 5). This is another place where our subtler understanding of rationality becomes important. The discussion of interpreting the choice function is an example of where this arises.

The same theme arises in chapter 6 as well: we shouldn't take evidence from climate models as giving us determinate answers to questions about the future evolution of the climate, but rather as giving us *some* indications of what sorts of scenarios to prepare for. The final chapter (chapter 7) again stresses this point that it is important to be clear what we can expect – in terms of decision relevant advice – from our modelling of the climate. Of course, this negative conclusion is not of much help to policy makers who actually have to make decisions based on this incomplete picture of the future evolution of the climate; the last chapter discusses some strategies for making decisions when the evidence doesn't give you enough advice.

These last two chapters should be seen as a case study of uncertainty in science that is motivated by the same concerns that led to the formal conclusions of the earlier chapters. That is, in both cases I am motivated by my interest in a weaker standard of rationality, and by a concern for adequate representation of uncertainty for the purposes relevant to the context.

2. A framework for belief, uncertainty, value and action

> The greatest challenge to any thinker is stating the problem, in a way that will allow a solution.
>
> *(Bertrand Russell)*

This chapter serves as a summary of the formalism used throughout the rest of the dissertation. It also serves as an introduction to the ideas and concepts appealed to. Ultimately, I am interested in *rational belief and action under severe uncertainty*. These concepts – "rational", "belief", "action", "severe uncertainty" – will all need explanation, as will some other related concepts.

I will first discuss modelling decision problems in the abstract. Then I will move on to a discussion of the formal details. I have split this discussion into two parts. These are the *objective* and *subjective* parts of decision problems. This distinction is not sharp, and it is a little arbitrary. The idea is that the objective aspects of a decision problem will be common to all agents. Agents may, however, differ in the subjective parts. The same acts will be available to all agents, but different agents might value the outcomes differently. The distinction is not sharp because sometimes the chances of the events determine the beliefs in them. The intuition behind drawing the distinction where I do is that there is nothing "normative" going on in the objective half: this is supposed to be just a neutral characterisation of the problem. The normativity comes in in the subjective part.[1]

[1] Bermúdez (2009) discusses the possibility of normatively assessing the way you have framed the decision problem. I don't assess that sort of thing here.

2.1. Modelling decision problems

When one is modelling a decision scenario, there are three kinds of things that are relevant. There are objects of value, objects of belief, and objects of choice. This is "object" in a grammatical sense: these are objects in that they are the object of a verb. In the current context, it is the decision maker who is the subject of those verbs. For example, the objects of value are the things the subject values. I don't mean to suggest that all objects of value are *material objects*.

The objects of value are the things you are concerned to bring about, or to prevent from happening. For example, "getting caught in the rain"; "winning the bet". If there were no such objects of value, then there'd be no reason to "decide" anything. The objects of belief are the contingencies that will affect what happens. They influence which of the objects of value end up occurring. For example, "it is raining"; "your horse wins the race". Finally, the objects of choice also affect which outcomes come about, but the objects of choice are those eventualities you have it in your power to control. For example "take the umbrella"; "bet on the horse called *Categorical Imperative*".

Let's take a simple example. In this example, the objects of value are "get wet" which is a contingency you are keen to avoid; "carry umbrella" which is a nuisance, but not so bad as being caught in the rain; and "don't get wet". The objects of belief are "it rains" or "it doesn't rain". The objects of choice are "bring umbrella" and "don't bring umbrella". It should be clear that both the objects of belief and the objects of choice influence which of the objects of value ends up coming about. It should also be clear that it is in your power to choose between the objects of choice – you have it in your power to bring about the fact that you take an umbrella or not – but you don't have the same power over the objects of belief (whether or not it rains).

	It rains	It doesn't rain
Bring umbrella	Don't get wet	Carry umbrella
Don't bring	Get wet	Don't get wet

Table 2.1.: A simple decision problem

Fixing the relevant elements of these three categories is the first part of modelling a decision problem. One might worry that these types of objects all seem to be somewhat "subjective". That is to say, it isn't really the proposition "it is raining" that influences whether or not you get wet, but rather, it is *what makes it true that it is raining* that causes the getting wet. It might be that "it is raining" logically entails "you get caught in the rain", but what you value is not the truth value of that proposition: you (dis)value *the actual state of getting caught in the rain*. This subtlety can be glossed over by assuming that the logical relationships between these subjective objects mirror the relationships[2] between the *things in the world* that determine your happiness. That is, the "world" makes the proposition true. The other way of sweeping this issue aside is to make sets of possible worlds be the objects of belief. I discuss this later, when I introduce events and truth.

How sharp this tripartite distinction between objects of value, belief and action is, is a matter of controversy. Savage (1972 [1954]) makes the separation of the categories absolute. Jeffrey (1983) makes these things less distinct. I return to this question later in section 2.4.2, once I have set out the formal details of the frameworks.

2.2. Objective elements of decision problems

Ultimately, we are interested in rational belief and rational action, but first we need to work out what beliefs *are*, and what action involves. Belief is a propositional attitude. It is important for a number reasons. There are norms that govern rational belief: directly through what it is reasonable to believe; and indirectly through belief's action-guiding role and what actions are reasonable. Before we get to belief, we need a characterisation of what the objects of belief are.

2.2.1. Events

Belief is belief *about* something: You can have beliefs about whether it will rain, about which horse will win the race and so on. When we model belief, we make the objects of belief sentences that say things about the world. A belief is an attitude directed at things in the world. The content of a belief can be something like "It is raining" or "The die lands on a four". These are sentences about states of affairs; about events in the world. Using these basic sentences we can construct compound sentences like "The die lands on a four **or** it is raining".[3] This "**or**" is formalised as "$\vee$". Using "X" to stand for "The die lands on a four" and "Y" to

[2] I leave it open for now whether these relationships are logical, causal, evidential...

[3] While it is possible to construct these logical compounds, I'm not sure how a decision problem would rest on the truth value of a disjunction with disjuncts from different domains like this.

stand for "It is raining", we can rewrite this compound as "$X \vee Y$". More generally, for any X_1, X_2 sentences, we can construct the compound $X_1 \vee X_2$. This is also a sentence, we can have beliefs about it. Of course, what beliefs we have about the compound will be somehow related to what beliefs we have in the basic sentences.

It is useful to have some other kinds of compound sentences available to us: let "$X \wedge Y$" mean the proposition "X **and** Y"; and let "$\neg X$" mean "**it is not the case that** X" or "**not** X" for short. Again, what beliefs you have in these compounds is related to your beliefs in the basic sentences. We can then write complex compound sentences like this: $(X \vee Y) \wedge \neg(\neg Z \wedge Y)$.

More formally, we have a set of elementary letters $L = \{x_1, x_2 \dots x_n\}$ which is finite unless otherwise stated. The infinite case introduces some complexity which is, for the most part, besides the point in the current project.[4] We have a set of sentences built up from these propositions and the usual logical connectives $\neg, \wedge, \vee$. Two important elements of this logic, the tautology and the contradiction, will be identified as $\top$ and $\bot$ respectively. The tautology is always true, the contradiction is never true.[5] This terminology follows Paris (1994). We can identify $\top$ with $X \vee \neg X$ and $\bot$ with $X \wedge \neg X$. Jeffrey's system is importantly different: Jeffrey's algebra is atomless. This means that, among other things, there are infinitely many elementary letters.[6]

One might want to disagree with the understanding of belief as a propositional attitude. "No," you might say, "my beliefs are *about the world*." The objects of belief are in the world. However, I think this way of speaking doesn't do justice to the intuition that beliefs are *intensional*. For example, you can believe that Clark Kent is in the room, but not believe that Superman is in the room: this is not contradictory. However, Clark Kent's being in the room is extensionally the same as Superman's being in the room. So it seems that if beliefs are about the world directly, you can't have the above pair of beliefs. That's not to say that you can't make sense of this aspect of intensionality and have your beliefs be about the world, but it's subtle. Weisberg (2011) makes the same point. Having the objects of belief be "in the world" also makes i somewhat mysterious what compounds like

[4] However, there are cases where it is *necessary* to work with an infinite language. Most notably, the theorems of Savage (3.2.3), Fine/Villegas (3.2.4) and Cox (3.4.1) don't work unless the language is infinite.

[5] More carefully, the tautology is true under all assignments of truth values to the elementary letters; the contradiction, false under all such assignments.

[6] Whether the atomlessness is a requirement of rationality, rather than a structural feature required for the representation theorem will be discussed later.

$X \vee Y$ correspond to. Are these logical operators to be understood as operating on the things in the world? What would that mean? It seems more straightforward to have the objects of belief be linguistic: we know how to handle logical operators on linguistic things.

The sentences that are the objects of belief are sentences *about the world*, so belief isn't totally divorced from the world. This intermediate step allows us to deal with intensionality in a natural way. That said, this property of my preferred way of understanding belief won't feature in this dissertation.

We are going to be discussing degrees of belief in these sentences. It is perhaps reasonable, and certainly convenient, to stipulate that logically equivalent sentences are believed equally strongly. So we will build this constraint into our framework by only defining our belief functions over equivalence classes of logically equivalent sentences. So take the set of compounds build up from L and the logical connectives, and then use the equivalence relation "$\equiv$" to partition this into classes of equivalent sentences.[7] We will use elements of the equivalence class to stand for the class. Call such equivalence classes *propositions*. This is called the "Lindenbaum algebra" and behaves well under the logical connectives. Call this SL.[8]

Real human reasoners typically don't believe all logically equivalent propositions equally strongly. I am uncertain of many mathematical expressions which are equivalent either to $\top$ or to $\bot$. But as long as I don't know which, I am uncertain. However, *logical omniscience* does still seem normatively compelling, even if it is unrealistic. What reason could you have for believing logically equivalent propositions to different degrees apart from logical ignorance? It does at least seem required of you that if you know P and Q are equivalent, then you should believe them to the same degree. I think Kyburg (1983) has it right when he says that belief in X involves a commitment to appropriate beliefs in things logically related to X. For example, Kyburg believes the axioms of set theory, and he thereby takes himself to be committed to all the theorems of set theory, even those he doesn't know. He might act against such a commitment through ignorance, but he should feel foolish if such a failure was pointed out to him. Full logical omniscience of the kind built in by the focus on the Lindenbaum algebra is obviously too strong a

[7] I haven't yet introduced the machinery to make precise exactly what this "$\equiv$" relation is, but I hope it is obvious from standard accounts of propositional logic. In fact, later I will discuss different sorts of equivalence relation, so this initial vagueness serves a purpose.

[8] Note that it follows that it doesn't matter what X we use to pick out $\top$ and $\bot$: they are all in the same equivalence class anyway.

requirement. Relaxing this assumption would be an interesting project, but it is not a project I shall undertake here.[9]

As we have seen, events have a structure: we can use logical operators like $\wedge, \vee, \neg$ on them. Specifically, they have the structure of an algebra. For our purposes, this basically means that if $X, Y \in SL$ then so are $X \wedge Y, X \vee Y$ and $\neg X$.

As hinted at earlier, there is an alternative way to cash out what an event is. This is to say that an event is a set of possible worlds. And then define beliefs as being about sets of possible worlds. This is the way Halpern (2003) starts. Then events have a kind of set-structure, rather than logical structure: $X \cup Y$ is an event that corresponds to $X \vee Y$ as described above. That is, it is an event which obtains whenever at least one of X or Y does. This way, which corresponds to the "extensional" understanding of belief alluded to earlier, doesn't allow the kind of Superman/Clark Kent beliefs, nor does it allow failures of logical omniscience. Again, logical omniscience and intensionality won't feature in this dissertation, I mention these subjects only to justify my choice of logical rather than set-theoretic framework. These two understandings of what the content of belief are are not really in conflict, anyway. Indeed, the translation between them is straightforward. But to discuss it, I need to introduce another new idea.

2.2.2. Truth

Believing true propositions is the aim of rational belief.[10] What propositions are true is what decides which actions lead to which outcomes. So let's look at truth. We have a truth valuation function $\mathbf{v}$ which outputs $\mathbf{v}(X) = 1$ if X is true and $\mathbf{v}(X) = 0$ if X is false. That is,

Definition 2.2.1 $\mathbf{v}\colon SL \to \{0, 1\}$ *is a (classical) truth valuation function if*

- $\mathbf{v}(X \vee Y) = \max\{\mathbf{v}(X), \mathbf{v}(Y)\}$
- $\mathbf{v}(X \wedge Y) = \min\{\mathbf{v}(X), \mathbf{v}(Y)\}$

[9] As a matter of fact, since all the belief functions I look at have the property of being monotonic, I don't *need* to stipulate this. It follows from monotonicity that logically equivalent sentences are believed equally strongly. However, I include this discussion of the Lindenbaum algebra here because I want to make it clear that this is a choice about how I am conceptualising the *objects of belief*. In the current setting, logically equivalent sentences being believed equally strongly is overdetermined: either of the Lindenbaum algebra move or monotonicity would be sufficient to secure the property.

[10] Or that is at least part of the aim of rational belief. Disbelieving false propositions is just as important. In any case, truth and falsity are inextricably linked concepts, duals of each other.

- $\mathbf{v}(\neg X) = 1 - \mathbf{v}(X)$

Let's call the set of such functions $\mathbf{V}$. Your intuitions about how truth interacts with the standard logical operators should allow you to see the reasonableness of these properties for functions in $\mathbf{V}$. $\mathbf{v}$ describes how the world is with respect to all the propositions of interest.

With this terminology in place, we can now see how to translate between sets of possible world talk and proposition talk. Since $\mathbf{v}$ assigns a 1 or a 0 to all members of SL, it effectively picks out a unique "possible world". Call $[X]$ the set of $\mathbf{v} \in \mathbf{V}$ such that $\mathbf{v}(X) = 1$. $[X]$ is the set of possible worlds where X is true. It should be clear that $[X \vee Y] = [X] \cup [Y]$ and $[X \wedge Y] = [X] \cap [Y]$. Also, $[\neg X] = [\top]\backslash[X]$. Finally, $[X] \subseteq [Y]$ if and only if $X \vDash Y$. Thus, talking of propositions or of sets of worlds amounts to the same thing.[11]

Sometimes I will borrow some "set talk" to talk about propositions. I will often use "A partition of SL" to mean "A set of sentences of SL such that every $\mathbf{v} \in \mathbf{V}$ makes exactly one of them true." I hope it's clear that if the X_i have this property, the $[X_i]$ do indeed partition the set $\mathbf{V}$. I will also, on occasion talk of a "subset" of Y. By this I mean an X such that $[X] \subseteq [Y]$. X is a "subset" of Y when Y is a consequence of X.

We can now properly define the equivalence relation at the basis of our Lindenbaum algebra as follows: $X \equiv Y$ iff $[X] = [Y]$. We are operating purely semantically here. This might seem a rather round-about way of doing things. Why not just start with beliefs being over sets of "possible worlds" and be done with it? This is a very common approach in the literature. The reason to take this long way round is that the sets of possible worlds approach builds in the assumption of logical omniscience at the most basic level, and it is thus hard to relax. By taking this circuitous route to the basic set up, I have left open the possibility of relaxing logical omniscience. That is, I can replace "$\equiv$" with "$\cong$", some alternative equivalence relation signifying "known mutual entailment" or something like that. As long as this "plays nicely" with the logical connectives,[12] then we can construct a quotient algebra in the same way we can construct the Lindenbaum algebra out of the "$\equiv$" relation.

If L contains n elementary letters $\{x_1, \dots, x_n\}$, then there are 2^n distinct propositions of the form: $\pm x_1 \wedge \pm x_2 \wedge \cdots \wedge \pm x_n$. Where $+x_i$ means x_i and $-x_i$ means

[11] Once we have in place the Lindenbaum algebra, that is.

[12] Meaning, if $X \cong X'$ and $Y \cong Y'$ then $X * Y \cong X' * Y'$ for any connective "$*$". In other words, "$\cong$" is a congruence relation. Thanks to David Makinson for helping with this point. Note that such an approach would require a weaker version of monotonicity as well.

$\neg x_i$. For every $\mathbf{v}_i$ there is exactly one such proposition that it makes true. Call it t_i. That is, $\mathbf{v}_i(t_i) = 1$ and for $j \neq i$, $\mathbf{v}_i(t_j) = 0$. The t_is partition SL. These t_i are the *atoms* of the algebra SL.

To recap, we have a set of basic elementary letters $L = \{x_1, x_2, \ldots, x_n\}$ and a collection of compounds of L made by using the connectives $\wedge, \vee, \neg$. We take equivalence classes of these things using the logical equivalence relation, we call this SL. These are the things we have beliefs over.

2.2.3. Outcomes

Another kind of thing of interest in the context of decision making is *outcomes*: consequences, objects of value. Outcomes are the kinds of thing that you value or disvalue; the kind of thing you have preferences over; the things that result from your choosing to act one way or another. For example "Eat ice cream" might be the outcome of choosing to have dessert. Savage (1972 [1954]) leaves outcomes as more or less structureless, uninterpreted primitives in his theory. They simply form a set; you have preferences over this set;[13] acts are functions which map into this set. That's all there is to say about them.

Jeffrey (1983) on the other hand, makes outcomes the same sort of thing as events (and everything else): outcomes are simply a type of proposition. They are a type of proposition that we have preferences over. Outcomes are propositions of the form $X \wedge A$ where X is a state of the world and A is an act. An outcome *just is* the conjunction of the act you perform and the state of the world. Understanding what an act is, is just understanding the conjunctions of the act with the possible propositions. Joyce (1999) points out that "Jeffrey outcomes" can't appear in more than one space in the decision matrix. So Joyce distinguishes fine-grained Jeffrey outcomes from coarse-grained Savage outcomes which are disjunctions of Jeffrey outcomes that you are indifferent between. In another sense, Savage outcomes are more fine-grained than Jeffrey outcomes. This is so since Savage outcomes must specify all the value-relevant details of the combination of act and event, whereas Jeffrey outcomes can leave many details underspecified. The relationship between these two concepts is subtle, and need not concern us here, since in what follows I will be mostly sticking to outcomes that are of fixed monetary value.

Anscombe and Aumann (1963) give some structure to their outcomes. Some outcomes are just basic goods like Savage's. But others are "lotteries". A lottery

[13] In fact, according to Savage, you only have preferences over them in virtue of having preferences over the constant acts, but *intuitively* outcomes are the things we have preferences over.

gives you one of a number of prizes with a certain probability. If the probabilities are known, it is called a *roulette lottery*. If the probabilities are unknown, it is called a *horse lottery*. The prizes can either be basic goods, or they can be more lotteries. The former sort are called *simple lotteries*, the latter *compound lotteries*. I won't discuss this addition much. For my purposes, it will suffice to take outcomes to be unstructured, unalloyed goods.[14]

If we effect the same liberalisation with outcomes as we did for events – namely to have differently described but logically equivalent outcomes treated differently – we can rationalise some kinds of experimentally observed preferences that don't fit the standard model.[15] Lois Lane can prefer being saved by Superman to being saved by Clark Kent, despite those outcomes being extensionally equivalent. Consider two different descriptions of extensionally the same outcome. Tversky and Kahneman (1986) discuss how people can have different preferences over these outcomes depending on how the outcome is framed. For instance, people respond differently to an act framed in terms of the number of deaths it would cause rather than the number of lives it would save, even if the prospects are extensionally the same. In effect, we allow the outcomes to be intensional in the same way we might allow the events to be intensionally described. This isn't a project I pursue here, but I note that it is possible.

The set of outcomes in a decision problem will be called $\mathbf{O}$. Outcomes are going to be described using $o_1, o_2 \dots$ and sets of outcomes shall be $O_1, O_2 \dots$ So $\mathbf{O} \supset O_1 \ni o_1$. This is a convention I shall try to roughly stick to with all my formalism. An "x" will be referred to by a small letter, a set of xs by a capital "X" and "the set of (all) the xs" by a bold capital "$\mathbf{X}$". Capital blackboard bold ($\mathbb{X}$) and script letters ($\mathcal{X}$) will then be used for other complex entities. Lowercase boldface will typically denote important types of function.

2.2.4. Acts

A third important aspect of decision problems is what the agent can *do* about it. If you are in a decision problem, you need to know what acts are available to you; what the objects of choice are. Savage thought of acts as functions from states to outcomes. In the current set up, this seem a little strange. Having acts be functions from propositions to outcomes seems to gloss over the fact that it is

[14]Later, we will see that I allow myself mixed acts. One can interpret mixed acts as being lotteries over basic acts. The difference is purely cosmetic.

[15]Bermúdez (2009, Chapter 3) discusses this possibility.

not the propositions that really determine which acts are successful. What makes taking your umbrella a successful act is that it is raining; not that the proposition "it is raining" is true. Despite this, I hope it is clear that this sort of approach will work fine, as long as what determines the truth of the propositions that are arguments of the act functions is completely distinct from the decision problem. In slogan form: "The world decides what propositions are true". Or perhaps "The truth of event propositions and the success conditions of acts have a common cause: the world". If some of the relevant propositions are in your control – or are correlated with propositions in your control – then lots of weird stuff can happen. To keep things simple, I demand that what act you perform cannot influence what event obtains. Bringing an umbrella cannot influence whether or not it rains. This is the condition of *act independence of states*: which state occurs is independent of which act is performed.

Jeffrey thought of acts as propositions you have it in your power to make true. To recover "Savage acts" in Jeffrey's framework, you have to enrich your language with a new kind of connective which is a kind of non-truth-functional conditional (Bradley 2007; Joyce 1999).

Acts are the things you choose among and which – along with the true state of the world – determine which outcome comes about. Given how outcomes were characterised, it seems clear that this is how things should be. Not having the clear distinction between the elements of the decision problem makes Jeffrey's theory harder to deal with, in some ways. Joyce follows Jeffrey in making acts be propositions. I will go against Jeffrey and side with Savage. Not because I think Jeffrey is wrong, but because I think that Savage's abstraction is suitable for my purposes. I discuss the differences between the frameworks in section 2.4.2.

The set of available acts is $\mathbf{A}$ and elements of $\mathbf{A}$ will be $a_1, a_2 \dots$ and sometimes $f, g \dots$ Typically the set of basic acts available will be finite. From these basic acts, other acts can be built up. If we have some kind of random device that, say, outputs a 1 with probability p and a 0 otherwise, we can generate mixed acts. $pf + (1-p)g$ is the act "get whatever f gets you with probability p, get whatever g gets you otherwise". Anscombe and Aumann, and von Neumann and Morgenstern allow themselves this kind of mixed act. There might be some subtlety attached to how mixed acts are generated, but I don't think this will impact on my current project (Schwarz ms.).

A paradigm example of an act is a bet or wager. I will have a lot more to say about bets when we get to the Dutch book argument in section 3.1. When you bet on a horse, or on your football team, you are betting on a proposition like "The horse will win" or "My team will win".[16] How much money you end up with depends on the eventual truth or falsity of that proposition. In the case of bets, it is clear what factors of the state of the world determine the success of the act. And thus, it is obvious which propositions serve as the proxies for those factors.

Acts can be defined as functions from propositions to outcomes or as functions from the atoms to outcomes or as functions from a partition of SL to outcomes.

2.3. Subjective elements of decision problems

We have now seen the "objective" parts of a decision problem: the events, the outcomes and the acts. As the agent facing the decision problem, you must bring with you the "subjective" parts of it: beliefs and values. These are the things that can be normatively assessed. It seems reasonable to say "Bob's beliefs are irrational" but it does not seem reasonable to say "Bob's options are irrational".[17]

There are two important things to consider when evaluating the available acts in a decision problem. The first is "how good are the possible outcomes?" and second "how likely is it that each of these outcomes will obtain?" Answering these questions in tandem gives you the concept of "expectation". How good do you expect an act to be? The goodness of an act will depend on what outcomes it leads to under what circumstances, and how likely those circumstances (events) are.

These ideas are encapsulated in several functions, which we shall discuss in turn. The basic functions are the belief function, which maps events to numbers; and the value function which maps outcomes to numbers. The derived function is the *expected value* function, which maps acts to numbers. We also look at some related concepts: preference, and choice functions.

For the time being, I am going to focus on belief functions, which are my main interest. I will, for the most part, just take it as read that there is a value function which behaves appropriately in whatever context I am discussing.

[16]For this to be a good example, it needs to be the case that you are otherwise uninterested in which outcome eventuates. That is, if you have a basic preference for your team to win, whether or not you have bet on them, then this isn't quite so straightforward.

[17]Of course, if you take the modelling of the decision problem to be a subjective process on the part of the agent, then it's reasonable to say that Bob has modelled the problem in an irrational way. But I hope my main point is clear. Bermúdez (2009) discusses rational assessment of the decision modelling procedure.

2.3.1. Relations and representation

Before moving on to the meat of this section, I want to review a couple of ideas that crop up in various places, and so are worth discussing briefly. The first is that of a *relation*. The second, that of a *representation*.

At various points we shall discuss a variety of different relations. One important kind of relation is the *preference relation* which relates outcomes, or sometimes acts. Say "$\varphi \succeq \psi$" means "φ is at least as good as ψ". Sometimes one considers a relation among events, with "$\varphi \succeq \psi$" meaning "φ is at least as likely as ψ". We have already met the entailment relation where "$\varphi \vDash \psi$" means "φ entails ψ".

Different kinds of relations will have different properties. What properties are appropriate for a relation depends on what the relation is for. Some properties that will be important in what follows are listed below.

Transitivity If $\varphi \succeq \psi$ and $\psi \succeq \xi$ then $\varphi \succeq \xi$

Completeness $\varphi \succeq \psi$ or $\psi \succeq \varphi$ for all φ, ψ

Weakly Connected If $\varphi \neq \psi$ then $\varphi \succeq \psi$ or $\psi \succeq \varphi$

Reflexivity $\varphi \succeq \varphi$ for all φ

Irreflexivity $\neg(\varphi \succeq \varphi)$ for all φ

Symmetry If $\varphi \succeq \psi$ then $\psi \succeq \varphi$

Asymmetry If $\varphi \succeq \psi$ then $\neg(\psi \succeq \varphi)$

Antisymmetry If $\varphi \succeq \psi$ and $\psi \succeq \varphi$ then $\varphi = \psi$

"$\succeq$" will be understood as a reflexive relation, while "$\succ$" will be the irreflexive part of it.[18] That is, we have the following two definitions.

Definition 2.3.1 *From the reflexive relation $\succeq$ define its irreflexive and symmetric parts:*

- *$\varphi \succ \psi$ iff $\varphi \succeq \psi$ and $\neg(\psi \succeq \varphi)$*
- *$\varphi \sim \psi$ iff $\varphi \succeq \psi$ and $\psi \succeq \varphi$*

Some trivial conclusions that we will need to appeal to later are collected in this lemma:

[18] Some of the relations in what follows will have subscripts: $\succeq_X$. The relations $\succ_X, \sim_X$ relate to $\succeq_X$ in the obvious way.

Lemma 2.3.1 *For reflexive* $\succeq$ *and* $\succ, \sim$ *defined as in Definition 2.3.1:*

- *If* $\varphi \succ \psi$ *then* $\neg(\psi \succ \varphi)$
- $\neg(\varphi \succ \varphi)$

Two important weakenings of transitivity should be mentioned here.

Acyclic For all $\varphi_1, \varphi_2, \dots, \varphi_n$ we have $\varphi_1 \succ \varphi_2, \varphi_2 \succ \varphi_3, \dots, \varphi_{n-1} \succ \varphi_n$ implies $\neg(\varphi_n \succ \varphi_1)$

Suzumura consistent For all $\varphi_1, \varphi_2, \dots, \varphi_n$ we have $\varphi_1 \succeq \varphi_2, \varphi_2 \succeq \varphi_3, \dots, \varphi_{n-1} \succeq \varphi_n$ implies $\neg(\varphi_n \succ \varphi_1)$

A transitive reflexive relation is known as a preorder. An antisymmetric preorder is a *partial order*. A set equipped with a partial order is known as a *poset*. A complete relation is sometimes called *total* or *connected*. A complete antisymmetric preorder is a *complete order*, *total order* or *linear order*. A complete, transitive, symmetric relation is called an *equivalence relation*. The relation of logical equivalence in our discussion of the Lindenbaum algebra is an equivalence relation in this sense.

One can think of a relation as a set of ordered pairs of elements in its domain. Call this set $R_{\succeq}$. We have $\varphi \succeq \psi$ if and only if $(\varphi, \psi) \in R_{\succeq}$. It then makes sense to talk about the intersection of relations, or their union in terms of the intersection (union) of the corresponding sets of ordered pairs. I will use this trick later. Note that $R_{\succeq} = R_{\equiv} \cup R_{\succ}$.

There are at least two places where we will have a relation over some domain (events, acts) and we will want to *represent* it by a function from that domain into another domain (typically the real numbers). We will see one example of this in the next section, but since this is something that will happen more than once, we discuss it briefly in abstract here.

The idea is that a function $f : \mathbb{X} \to \mathbb{R}$ *represents* a relation $\succeq_f$ if and only if:

$$\text{If } x \succeq_f x' \text{ then } f(x) \geqslant f(x') \tag{2.1}$$

For example, if the objects were sticks of various lengths, and the relation of interest was "is longer than" then the function that returns the length of the stick in centimetres *represents* the structure in the right way. That is x is longer than x' iff $\mathbf{len}(x) > \mathbf{len}(x')$. I will return to this length example in detail in section 3.2.1.

One can generalise this to representations where f is a function into an arbitrary space equipped with its own relation, but in general we will stick to representing things with real numbers. Chu and Halpern (2004, 2008) provide a *very* general theory of decision with representations by arbitrary posets, rather than the reals. For representation more generally, the classic work is Krantz et al. (1971).

2.3.2. Representing belief

From qualitative probability to belief functions

There are some things you believe more strongly than others. For example, I currently have a stronger belief in "It will rain tomorrow" than I do in "I will win the lottery this week". So there is a relation of *qualitative probability* or *strength of belief* between propositions in SL. The idea is that there is some relation $\succeq_{\mathbf{b}}$ over SL where "$X \succeq_{\mathbf{b}} Y$" is understood to mean "X is at least as likely as Y". We might proceed by considering what properties this relation of qualitative belief should have, but I will postpone that project until section 3.2.4. Typically, one understands degree of belief as being characterised by the *belief functions* that represent this relation.

Definition 2.3.2 *A* belief function *is a function* $\mathbf{b}\colon SL \to \mathbb{R}$.

Let $\mathbf{B}$ be the set of all such belief functions. Belief comes in degrees and belief functions are supposed to capture this notion of strength of belief. A higher number assigned to an event should be interpreted as a stronger belief. We've already seen one such kind of function: $\mathbf{V}$. We won't typically be thinking of the valuation functions as belief functions, but *they do satisfy the basic definition*.

Consider $\mathbf{b}(\top)$ and $\mathbf{b}(\bot)$ for some arbitrary $\mathbf{b} \in \mathbf{B}$. It seems reasonable that no proposition is more strongly believed than a tautology is believed. And no proposition is less believed than a contradiction. So it makes sense to restrict our attention to functions with the following property:

Bounded *For all X we have* $\mathbf{b}(\bot) \leqslant \mathbf{b}(X) \leqslant \mathbf{b}(\top)$

Now it also makes sense to fix the values of $\mathbf{b}(\bot)$ and $\mathbf{b}(\top)$. The actual value of the numbers is not important, only the relations among them (which encode the relations of relative strength of belief among the propositions). So let's fix these values:

Normalised $\mathbf{b}(\bot) = 0 \quad \mathbf{b}(\top) = 1$

Probability

The most important kind of belief function, the probability function, satisfies both of the above properties, and another important property.

Additive $\mathbf{b}(X \vee Y) + \mathbf{b}(X \wedge Y) = \mathbf{b}(X) + \mathbf{b}(Y)$

This isn't typically how you'll find additivity defined, but it's my favourite way. It is from Joyce (2009).

Definition 2.3.3 *A probability measure* $\mathbf{pr} \in \mathbf{B}$ *is a belief function that is* Bounded, Normalised *and* Additive

This is my "official definition" of probability. Note that probabilities so defined are finitely additive, but not necessarily countably additive.[19] Throughout this dissertation, I will restrict myself to finite additivity. There is some controversy over whether countable additivity is rationally required of belief, but that is a concern that does not have a bearing on the current project.[20]

The more standard way to define probability, I call the "incompatible propositions" definition. A probability measure $\mathbf{pr} \in \mathbf{B}$ is a belief function that satisfies these properties:

i$_I$ For all X we have $\mathbf{pr}(X) \geqslant 0$

ii$_I$ For all X, Y incompatible propositions, $\mathbf{pr}(X \vee Y) = \mathbf{pr}(X) + \mathbf{pr}(Y)$

iii$_I$ $\mathbf{pr}(\top) = 1$

Note that by induction, (ii$_I$) entails that for incompatible $X_1, X_2 \ldots X_n$, we have $\mathbf{pr}(\bigvee_{i=1}^n X_i) = \sum^n \mathbf{pr}(X_i)$.

Yet another way of defining probability – the "partition additivity" approach – is to say that $\mathbf{pr} \in \mathbf{B}$ is a probability if:

i$_P$ For all X we have $\mathbf{pr}(X) \geqslant 0$

[19] Countable additivity is the property that countable infinite disjunctions of propositions are equal to the sum of their individual beliefs.

[20] In fact, Williamson (2001) shows that any finitely additive probability function over a logical language has a unique extension to a countably additive probability function. So in the current setting, the only question is whether or not we allow infinitary disjunctions to be objects of belief. But the general point stands: one might want to consider beliefs over slightly different logic structures – structures for which Williamson's result does not hold. In these circumstances there would be a genuine question as to whether countable additivity was a requirement of rationality.

ii$_P$ If X_i partition SL then $\sum \mathbf{pr}(X_i) = 1$

These three ways of defining probability amount to the same thing.

Theorem 2.3.1 *The following things are equivalent:*

- *A belief function* $\mathbf{pr} \in \mathbf{B}$ *is* Bounded, Normalised *and* Additive*:*
- $\mathbf{pr}$ *satisfies the incompatible propositions definition*
- $\mathbf{pr}$ *satisfies the partition additivity definition*

Proof Let $\mathbf{pr} \in \mathbf{B}$ be a probability function as defined by the official definition. I show it also satisfies the incompatible propositions definition. (i$_I$) and (iii$_I$) follow straightforwardly from $\mathbf{pr}$ being Bounded and Normalised. If X and Y are incompatible, then $X \wedge Y \equiv \bot$, so $\mathbf{pr}(X \wedge Y) = 0$. So, by Additivity: $\mathbf{pr}(X \vee Y) + 0 = \mathbf{pr}(X) + \mathbf{pr}(Y)$. So (ii$_I$) is satisfied.

Next I show that the incompatible propositions definition implies the partition additivity definition. (i$_P$) is the same as (i$_I$). The elements of a partition X_i are such that their disjunction is $\top$. So by (iii$_I$) $\mathbf{pr}(\bigvee X_i) = 1$. Elements of a partition are incompatible, so by (ii$_I$) we know that $1 = \mathbf{pr}(\bigvee X_i) = \sum \mathbf{pr}(X_i)$.

Once we have shown that the partition additivity definition entails our official definition, we will be done. Let $\mathbf{pr} \in \mathbf{B}$ now satisfy the partition definition. For incompatible X, Y we have $\{X, Y, \neg(X \vee Y)\}$ is a partition. So is $\{X \vee Y, \neg(X \vee Y)\}$. From this it follows that $\mathbf{pr}(X) + \mathbf{pr}(Y) = \mathbf{pr}(X \vee Y)$. Now take X', Y' which are not necessarily incompatible. We have $X' \wedge \neg Y'$, $X' \wedge Y'$ and $\neg X' \wedge Y'$ are incompatible, and their disjunction is $X' \vee Y'$. So:

$$\begin{aligned}\mathbf{pr}(X') + \mathbf{pr}(Y') &= \mathbf{pr}(X' \wedge Y') + \mathbf{pr}(X' \wedge \neg Y') + \mathbf{pr}(Y' \wedge X') + \mathbf{pr}(Y' \wedge \neg X') \\ &= \mathbf{pr}(X' \wedge Y') + \mathbf{pr}(X' \vee Y')\end{aligned}$$

So $\mathbf{pr}$ satisfies our official definition of additivity. Since $\mathbf{pr}$ is nonnegative, and every proposition can be a member of some partition, boundedness and normalisation are also satisfied. ∎

We turn now to a relationship between truth and belief. This relationship will be important later.

Definition 2.3.4 *Let* $\mathbf{V}^+$ *be the convex hull of* $\mathbf{V}$. *That is,* $\mathbf{V}^+$ *is the set of* $\mathbf{b} \in \mathbf{B}$ *such that there exist* λ_i *with the following properties:*

$$\begin{aligned} \lambda_i &\geqslant 0 \\ \sum \lambda_i &= 1 \\ \mathbf{b}(X) &= \sum \lambda_i \mathbf{v}_i(X) \textit{ for all } X \end{aligned}$$

That is, $\mathbf{b}$ *is a weighted sum of truth valuations.*

With this definition, we can state a theorem that shows that $\mathbf{V}^+$ is exactly the set of probability functions in $\mathbf{B}$ when SL is finite. This result is apparently due to De Finetti, though I haven't found it in his *Theory of Probability*. Paris (1994) contains something close to this result on pp. 13–14, but it isn't as general as I'd like. That is, it only applies to classical truth value functions, and only applies to spaces that contain atoms. This is more or less the result I prove below. Paris mentions a different result due to Glenn Shafer, but gives no reference in Paris (2005 [2001]). Whether this also holds for infinite spaces I am not sure.

Theorem 2.3.2 $\mathbf{V}^+$ *is the set of all and only the probability functions over* SL.

Proof This proof sketch follows Paris (1994). First we show that if $\mathbf{pr}$ is a probability function, then there exist λ_i such that $\mathbf{pr}(X) = \sum \lambda_i \mathbf{v}_i(X)$ with $\sum \lambda_i = 1$. Recall, that for each atom t_i there exists exactly one $\mathbf{v}_i$ with $\mathbf{v}_i(t_i) = 1$. That is, exactly one valuation function makes true exactly one maximally specific of the ways the world could be. Let $\lambda_i = \mathbf{pr}(t_i)$. Since $\mathbf{pr}$ is bounded and normalised, $\lambda_i \geqslant 0$.

Now we must show that $\sum \lambda_i = 1$ and that for every $X \in SL$, $\mathbf{pr}(X) = \sum \lambda_i \mathbf{v}_i(X)$. Every X is a disjunction of t_is. $\sum \lambda_i \mathbf{v}_i(X) = \sum \mathbf{pr}(t_i)\mathbf{v}_i(X)$. Now the $\mathbf{v}_i(X)$ in the sum means that we effectively only count those t_i that are true when X is true. So the sum becomes the sum of the $\mathbf{pr}(t_i)$s where t_i is a disjunct of X. Thus $\sum \mathbf{pr}(t_i)\mathbf{v}_i(X) = \mathbf{pr}(X)$.

We now prove the other direction: given λ_i, we show that $\mathbf{b}(X) = \sum \lambda_i \mathbf{v}_i(X)$ is a probability function. $\mathbf{b}(X) \geqslant 0$ since all the λ_is are non-negative, and so are the outputs of $\mathbf{v}_i$. Take some partition of SL, call it $\{X_j\}$. To prove $\mathbf{b}$ is a probability function, all we need to do now is show that $\sum \mathbf{b}(X_j) = 1$ (See Theorem 2.3.1).

$$\begin{aligned} \sum_j \mathbf{b}(X_j) &= \sum_j \sum_i \lambda_i \mathbf{v}_i(X_j) \\ &= \sum_i \lambda_i \sum_j \mathbf{v}_i(X_j) \end{aligned}$$

Now since X_j partition SL, for any $\mathbf{v} \in \mathbf{V}$ we have $\sum_j \mathbf{v}(X_j) = 1$. Thus:

$$\begin{aligned}\sum_i \lambda_i \sum_j \mathbf{v}_i(X_J) &= \sum \lambda_i \\ &= 1\end{aligned}$$

■

This theorem will be useful later. For example, it is the basis of Joyce's argument discussed in section 3.3. Williams (2012a,b) discusses modified versions of this theorem for convex hulls of different kinds of sets of nonclassical valuation functions. For a set of nonclassical truth valuations $\mathbf{W}$, we can ask how to characterise the set of convex combinations of $\mathbf{W}$. Call it $\mathbf{W}^+$. For some particular sets $\mathbf{W}$, we know what the convex hull contains, but there doesn't seem to be a general result.

Conditional probability

Before moving on, I'd like to briefly consider an important modification to the probability framework. Instead of taking these *unconditional probabilities* $\mathbf{pr}(X)$ as the basic entity, we might consider taking our probability functions to be two-placed *conditional probability* functions $\mathbf{pr}(X|Y)$. We read this as "the probability of X *given* Y". We often interpret the conditional probability of X given Y as what your belief in X would be if you learned Y was true. In this section I talk as if this link between conditional probability and updated probability is uncontroversial. It is not, but it is intuitive, and it makes the presentation slightly less awkward. One could replace "learning X" with "supposing that X" throughout this section. Before we can define conditional probability, we need a slightly more sophisticated concept of the space of belief. My presentation follows that of Halpern (2003, p. 74–6). Instead of being defined over the set of sentences SL, conditional probability functions are defined over $SL \times SL'$ where SL' is a subset of SL that is closed under consequence. That is, if $X \in SL'$ and $X \models Y$ then $Y \in SL'$. Often SL' is taken to be $SL \backslash \{\bot\}$. A conditional probability function is a function $\mathbf{pr}\colon SL \times SL' \to \mathbb{R}$. Conditional probability measures satisfy versions of Bounded, Normalised and Additive for all $X, X_1, X_2 \in SL$ and $Y \in SL'$

Bounded $\mathbf{pr}(\bot|Y) \leqslant \mathbf{pr}(X|Y) \leqslant \mathbf{pr}(\top|Y)$

Normalised $\mathbf{pr}(\bot|Y) = 0$ and $\mathbf{pr}(\top|Y) = 1$ and $\mathbf{pr}(Y|Y) = 1$

Additive $\mathbf{pr}(X_1|Y) + \mathbf{pr}(X_2|Y) = \mathbf{pr}(X_1 \wedge X_2|Y) + \mathbf{pr}(X_1 \vee X_2|Y)$

These are basically the unconditional properties rewritten, replacing every instance of $\mathbf{pr}(\cdot)$ with $\mathbf{pr}(\cdot|Y)$ and they hold for all $Y \in SL'$. The only addition is the third component of the NORMALISED axiom which says that the conditional probability of Y given Y is also 1. This is plausible: on learning Y, you should come to believe Y to the highest degree. This isn't so much a substantive assumption as a characterisation of the kind of learning we are concerned with. So for each Y the function $\mathbf{pr}(\cdot|Y)$ is an unconditional probability function. In particular, it is common to identify the unconditional probability $\mathbf{pr}(\cdot)$ with the conditional probability function conditioned on $\top$: $\mathbf{pr}(\cdot) = \mathbf{pr}(\cdot|\top)$ Conditional probabilities also have a new axiom that describes how the first and second arguments interact:

CONDITIONAL *If $X_2, X_3 \in SL'$ and $X_1 \in SL$ then:*

$$\mathbf{pr}(X_1 \wedge X_2|X_3) = \mathbf{pr}(X_1|X_2 \wedge X_3)\,\mathbf{pr}(X_2|X_3)$$

Using the convention that $\mathbf{pr}(\cdot) = \mathbf{pr}(\cdot|\top)$, and setting $X_1 = X, X_2 = Y, X_3 = \top$ we can rearrange the terms in the above equation to give us the following:

$$\mathbf{pr}(X|Y) = \frac{\mathbf{pr}(X \wedge Y)}{\mathbf{pr}(Y)} \tag{2.2}$$

This equation is typically understood as the *definition* of conditional probability out of basic unconditional probabilities. On the current view, this equation is a *consequence* of the definition of basic conditional probabilities. This approach has the advantage that conditional probabilities can be defined even if the conditioning event – Y in the above example – has $\mathbf{pr}(Y) = 0$. The *ratio analysis*, as Hájek (2003) calls it, is undefined if the denominator is zero. Taking SL' to be the set of $X \in SL$ such that $\mathbf{pr}(X) > 0$, we have a perfect agreement between the two definitions of probability. The advantage of taking conditional probabilities to be the basic units is that it allows us to define larger SL' and still have everything work nicely.

Halpern offers a nice argument for why the ratio formula should characterise a kind of learning.[21] Given that this formula is a simple manipulation of CONDITIONAL, this argument lends weight to the reasonableness of that axiom. I will describe the theorem in words, and then give it formally. First, if you learn X, it seems reasonable that you now assign $\neg X$ minimum belief. This is basically just a characterisation of what kind of learning we are talking about: you learn

[21] The argument works just as well if you think of it as an argument about a certain kind of *supposing*.

X with certainty. Second, learning X with certainty (and nothing else) shouldn't change your beliefs with respect to how likely the various realisations of X are. Imagine you learn that your horse *Categorical Imperative* has placed: has finished in the top three. This information shouldn't change your attitudes to the relative likelihoods that the horse came first or second or third. That is, it would be strange if you thought it much more likely that the horse would come third than it is that it would come second, but on learning *only* that he placed, you become more confident of a second place relative to a third place. Halpern – using the "possible worlds" talk we discussed above – put the intuition this way:

> One reasonable intuition is that if all that the agent has learned is X, the the relative likelihood of worlds in X should remain unchanged.
>
> Halpern (2003, p. 72, with notational changes)

This is a condition of *rigidity*. The following theorem proves that any updated probability measure that accords with the above intuitions will satisfy the ratio formula with respect to the prior probability.

Theorem 2.3.3 *If the following properties hold for $X_1, X_2 \in SL$ and $Y \in SL'$ with $X_1 \models Y$, $X_2 \models Y$ and $\mathbf{pr}(X_2) > 0$:*

$$\mathbf{pr}(\neg Y|Y) = 0$$

$$\frac{\mathbf{pr}(X_1)}{\mathbf{pr}(X_2)} = \frac{\mathbf{pr}(X_1|Y)}{\mathbf{pr}(X_2|Y)}$$

then for any $Z \in SL$:

$$\mathbf{pr}(Z|Y) = \frac{\mathbf{pr}(Z \wedge Y)}{\mathbf{pr}(Y)}$$

Proof The proof can be found in Halpern (2003, p. 72) ■

I think that Hájek (2003, 2007) is right that conditional probabilities are the better basic building block. However, for my purposes the distinction will not matter. I will therefore simplify the formalism by sticking to unconditional probability functions. But it should be borne in mind that these are "really" conditional probabilities conditioned on $\top$.

For the alternative belief functions that I discuss below, there may well be conditional counterparts to the unconditional ones I discuss, but I shall not discuss them: Interesting as they may be, and foundationally superior as they may be, their interest is orthogonal to mine. I thus dispense with the extra formalism required to treat them and remain resolutely unconditional.

Capacities

In characterising probability theory above, I mentioned a few properties of belief functions. I think that (Bounded) is non-negotiable. I also think that (Normalised) is effectively the articulation of a convention, rather than a substantive restriction on belief functions. So all the alternative belief functions will satisfy these properties. I call these the basic conditions.

Another property that I feel is pretty important is the following:

Monotonic *If* $X \models Y$ *then* $\mathbf{b}(X) \leqslant \mathbf{b}(Y)$

Note however that anyone committing the conjunction fallacy violates this property.[22] Non-monotonic reasoning is a big research area in logic, and an important one. But that is about *inference* whereas this property is about static beliefs at a time. I think it has some intuitive weight.

Our first, most permissive set of belief functions is the Choquet capacity:

Definition 2.3.5 $\mathbf{cq} \in \mathbf{B}$ *is a* Choquet capacity *if it is* Bounded, Normalised *and* Monotonic.

See Eichenberger and Kelsey (2009) for a discussion of Choquet capacities. Call **CQ** the set of all Choquet capacities. This is quite a permissive framework. Indeed, for most purposes, Choquet capacities will not be informative enough. The (Monotonic) property is quite a weak property. It only tells us, for instance, that $\mathbf{b}(X \vee Y) \geqslant \mathbf{b}(X)$. It doesn't tell us how $\mathbf{b}(X \vee Y)$ relates to, say, the sum of $\mathbf{b}(X)$ and $\mathbf{b}(Y)$. This motivates our next definition:

Definition 2.3.6 *A belief function* $\mathbf{b} \in \mathbf{B}$ *is* 2-monotone *if:*

$$\mathbf{b}(X \vee Y) \geqslant \mathbf{b}(X) + \mathbf{b}(Y) - \mathbf{b}(X \wedge Y)$$

A 2-monotone capacity is (Monotonic).[23] Why this property is called "2-monotone" will become clear once we introduce the more general *n-monotone* property. Before

[22] Since $X \wedge Y \models X$, we must have $\mathbf{b}(X \wedge Y) \leqslant \mathbf{b}(X)$. But this is exactly what *doesn't* happen with people committing the conjunction fallacy: they take "Linda is a bank teller" to be less probable than "Linda is a bank teller and an active feminist." See Tversky and Kahneman (1983) for more on the conjunction fallacy.

[23] If $X \models Y$ then $Y \equiv X \vee (\neg X \wedge Y)$. So $\mathbf{b}(Y) = \mathbf{b}(X \vee (\neg X \wedge Y) \geqslant \mathbf{b}(X) + \mathbf{b}(\neg X \wedge Y) - \mathbf{b}(X \wedge \neg X \wedge Y)$. So $\mathbf{b}(Y) \geqslant \mathbf{b}(X) + \mathbf{b}(\neg X \wedge Y)$. The result follows since $\mathbf{b}$ is nonnegative.

we do that, let's have a look at one more instance of it: 3-monotone:

$$\begin{aligned}\mathbf{b}(X \vee Y \vee Z) \geqslant \mathbf{b}(X) &+ \mathbf{b}(Y) + \mathbf{b}(Z) \\ &- \mathbf{b}(X \wedge Y) - \mathbf{b}(X \wedge Z) - \mathbf{b}(Y \wedge Z) \\ &+ \mathbf{b}(X \wedge Y \wedge Z)\end{aligned}$$

The degree of belief in the disjunction is the sum of the singleton beliefs, minus the sum of the beliefs about pairs, plus the belief in the triple conjunction. n-monotone just extends this basic *inclusion-exclusion* idea to arbitrary n. The idea is simple, its actual description takes some getting used to.

Definition 2.3.7 *A belief function* $\mathbf{b} \in \mathbf{B}$ *is* n-monotone *if:*

$$\mathbf{b}\left(\bigvee_{i=1}^{n} X_i\right) \geqslant \sum_{i=1}^{n} \sum_{\{I \subseteq \{1,\dots,n\} : |I| = i\}} (-1)^{i+1}\, \mathbf{b}\left(\bigwedge_{j \in I} X_j\right)$$

This is basically an "inclusion-exclusion" rule. It says that belief in n-element disjunctions is related to sums of belief values of conjunctions of subsets of the big disjunction. Taking each part of the right hand side from the inside out:

$$\mathbf{b}\left(\bigwedge_{j \in I} X_j\right)$$

I is some subset of $\{1,\dots,n\}$ so this is the belief in the conjunction of some subset of disjuncts of our big disjunction.

$$(-1)^{i+1} \quad \mathbf{b}\left(\bigwedge_{j \in I} X_j\right)$$

The $(-1)^{i+1}$ means we add conjunctions with an odd number of elements, and subtract the even ones.

$$\sum_{\{I \subset \{1,\dots,n\} : |I| = i\}} (-1)^{i+1} \quad \mathbf{b}\left(\bigwedge_{j \in I} X_j\right)$$

I ranges over all i-element subsets of our big disjunction. So we add or subtract all the i element conjuncts, depending on whether i is odd or even.

$$\sum_{i}^{n} \sum_{\{I \subset \{1,\dots,n\} : |I| = i\}} (-1)^{i+1} \quad \mathbf{b}\left(\bigwedge_{j \in I} X_j\right)$$

Finally we sum over the size of the subsets from 1 to n.

In the 2-monotone case this amounts to adding the singletons and subtracting the conjunction of the two. 3-monotonicity amounts to adding the singletons, subtracting all the conjunctions or pairs and then adding the conjunction of all three. An n-monotone function is $(n-1)$-monotone.[24]

Dempster-Shafer belief

Now we can introduce *Dempster-Shafer belief functions*. DS Belief functions will offer a middle ground between the informative, restrictive probabilities, and the looser Choquet framework.

DEFINITION 2.3.8 *A* Dempster-Shafer belief function *is an infinite-monotone capacity.*

A DS belief function has the n-monotone property for all n.

A useful way to think about DS belief is in terms of the associated concept of a mass function.

DEFINITION 2.3.9 *Call $m(\bullet)$ a* mass function *when:*

$$m(\bot) = 0$$
$$m(X) \geqslant 0$$
$$\sum_{X \in SL} m(X) = 1$$

Note that a mass function *is not* a representation of belief in the same way that **b** is. That is, more strongly believed events needn't have more mass. A mass function characterises "how much belief to associate with X, that can't be associated with subsets of X". A mass function represents belief through the associated belief function.

With this concept of a mass function we can characterise DS belief in a different way. We also characterise the related concept of a plausibility function **plaus**.

DEFINITION 2.3.10 *Call* **bel**(X) *a D-S belief function and* **plaus**(X) *a D-S plausibility function if, for some mass function m:*

$$\mathbf{bel}(X) = \sum_{Y \models X} m(Y) \qquad \mathbf{plaus}(X) = \sum_{Y \not\models \neg X} m(Y)$$

[24] Add $\bot$ as a final disjunct and it all cancels out.

The belief, or support, of X is the sum of the masses of the subsets of X. If the mass function is non-zero for atoms only, then the belief function is a probability measure.

The DS belief function **bel** is interpreted as the extent to which the evidence supports the conclusion X. The plausibility function **plaus** represents the extent to which the evidence *fails to support the contrary hypothesis*. In other words, **plaus** represents the extent to which the evidence is consistent with X.

Call **DS** the set of all DS belief functions. It is clear that, for all X we have $\mathbf{plaus}(X) = 1 - \mathbf{bel}(\neg X)$, so **DS** adequately captures both functions.

Theorem 2.4.3 of Halpern (2003, p. 36) shows that the two formulations of DS belief functions are effectively interchangeable in finite domains. That is, whether you take a DS function defined as an infinite-monotone capacity or defined by its associated mass function amounts to the same thing, at least for finite domains. For infinite domains, there are belief functions with no corresponding mass function.

These classes of belief functions are not unrelated. Every probability function is a D-S belief function, and every D-S belief function is a Choquet capacity. In other words: $\mathbf{V} \subset \mathbf{V}^{+} \subset \mathbf{DS} \subset \mathbf{CQ} \subset \mathbf{B}$.

2.3.3. Representing value

Preference relations

We need a way to express an agent's having a preference between things. We take $\varphi \succeq \psi$ to mean "φ is at least as good as ψ" and $\varphi \succ \psi$ to mean "φ is better than ψ". Call these weak preference and strong preference, respectively.

In section 3.1, the preference relation will be among bets, but in section 3.2 the preference will be among acts more generally. Elsewhere, one might wish to consider a preference relation among outcomes.

Value functions

As in the case of qualitative probability and belief functions, preference can be represented by value functions or utility functions. The idea is the same as for representing belief. The preference relation "$\succeq_u$" is represented by a value function, or utility function u iff: whenever $o \succeq_u o'$ we have $u(o) \geqslant u(o')$.

We extend this u function on outcomes to an "expected value" function on *acts* as follows. The idea is that in making a decision you weight the values you associate with the outcomes by how likely you consider those outcomes to be. As we have seen, what event obtains influences what outcome occurs. So the weights on the outcomes should connect with the degrees of belief in the events, somehow.

In the standard Savage framework, this amounts to multiplying the utility of the outcome which is the result of the act function acting on an event, by the probability of that event, and summing over the possible events. That is, we find some partition of SL, $\{X_i\}$ such that the acts we are considering are constant on each X_i and we sum over these events.

Definition 2.3.11 *The* expectation *of an act a with respect to a probability function* **pr** *is:*

$$E_{\mathbf{pr}}(a) = \sum_i \mathbf{pr}(X_i)u(a(X_i))$$

Choice of partition isn't that important here, since any refinement of the partition will give you the same value. The same is not true for other kinds of belief function. This problem will be taken up later.

The time spent on representations of belief versus representations of value is indicative of my concerns in the rest of the dissertation. For the most part, I take it for granted that the value of an outcome is adequately represented by the cash value associated with it.

2.4. Difficulties, caveats, alternatives

This section collects a miscellany of things I feel I should mention before moving on to the more substantive chapters.

2.4.1. Some standard caveats

Before continuing, we need to make clear a couple more caveats. These are fairly standard simplifying assumptions in the literature and nothing hangs on these details. We assume that you are risk neutral. This means, for instance that a preference among bets reflects your opinion of the relative value of the bets rather than a preference to avoid taking risks. (See Allingham 2002, Chapter 4; Binmore 2008, pp. 44–8; Peterson 2009, pp. 179–83). For example consider flipping a fair coin. Risk neutrality would mean that you would be indifferent between the following two bets: (a) paying one Euro for the chance to win two Euros if Heads comes up (b) paying one million Euros for the chance to win two million Euros if Heads comes up. Most people are happier to take the first rather than the second bet. In general, people are sensitive to the variance of a bet, not just its mean. This illustrates that people are not normally risk-neutral, but that they are risk-averse: they tend to avoid big risks. In general people are more wary of bigger gambles, even if the potential winnings are proportionally bigger as well. In fact, conceptually, there are two distinct things going on here – even if in practice they are dealt with formally in the same way. First, people tend to disvalue big risks: gambles with high variance in expectation are dispreferred. Second, the value of one Euro depends on how much money you already have: A million Euros is not a million times as good as one Euro. In both cases the standard formal assumption is that utility is linear with money. This builds in both that risk has no disutility, and also that money does not have diminishing marginal value. I am somewhat unhappy with the conceptual identification of these things that is often made, but for the present, I am happy to make both assumptions. I am not endorsing the claim that risk neutrality is rational. Indeed, I think there may be good reasons for thinking it isn't, except in very special cases. Okasha (2007) gives an evolutionary argument for the rationality of some level of risk aversion. Buchak (ms.) sets out a decision theory that allows a level of risk aversion to be rational. I think these are interesting projects, but I think I can safely bracket these issues given my focus.

We also assume a "state independence". That is, how good the outcome is does not depend on whether you have bet on it or not. The sort of things we want to rule out are the following: imagine you are betting on the weather and the possible prizes include an umbrella, sunglasses and so on. Now, which weather event occurs affects your attitude toward the prizes: presumably you value the umbrella much more if it is raining versus its being sunny, and vice versa for the sunglasses. This makes things messy, so let's stipulate that the value of the prizes is independent of the state that is actualised. I discuss state independence in more depth when I deal with representation theorems in section 3.2.

There are some further simplifications I shall be making. The decisions I consider take place *at a time*. So there are no *sequences* of decisions. I also keep your utility function determinate: I will not allow imprecise utilities for the time being.[25] I will also avoid the issue of whether sets of probability ought to be convex, I don't think they need be, although in this dissertation I normally assume they are.[26]

I am not arguing that these caveats are always reasonable to impose. But I think

[25]But see Bradley (2009)

[26]But see Kyburg and Pittarelli (1992, §4) for some discussion of problems with convexity. See also Bradley (2009) for comments on convexity when both probabilities and utilities are imprecise.

they are a useful abstraction, and the problems arising from relaxing them are orthogonal to the problems I am focusing on here.

2.4.2. Jeffrey's system, computer architecture and abstraction

It may seem strange that Jeffrey makes everything into propositions: intuitively, acts, states and outcomes are all rather different kinds of things. I want to explain why Jeffrey's approach makes sense, and then explain why I don't follow him.

Think about an integrated circuit. Everything is just a number:[27] the machine code instructions, the data, the addresses of the data... But, what makes things work is the *way* the number is treated: that's what makes something a machine code, or an address or what have you. A number in memory is treated as an instruction if, when it is read, it leads to particular transistors to turn particular circuits on or off. What makes some number a piece of data is if it is used to decide whether to send a high or a low signal through those circuits controlled by the instruction. What makes a number a memory address is if the instruction that preceded it is the sort of instruction that takes an address as an argument.[28]

In the same way, despite acts, outcomes and states all being the same sort of things, they are dealt with differently. Indeed, one might think of Savage's approach as formalising the separation that is implicit in the attitudes we have to the various kinds of propositions. To stretch the analogy further, higher level computer programming languages will abstract away from the machine code. This involves distinguishing different kinds of entities (often called *types*). So despite the fact that everything is fundamentally a number, some kinds of symbols will be used to refer to instructions, some other method will be used for referencing addresses, and some other way to indicate an actual number. Higher level programming languages will have different kinds of objects: strings, floating point numbers, signed integers, unsigned integers, lists... These are all, at base, numbers in memory, but it's useful to have different shortcut methods for manipulating them, since they need to be treated differently. If two numbers are representing particular strings of letters, then it doesn't make sense to add them together. You can do it, but the result doesn't have any meaningful interpretation.[29]

[27] Or not even: everything is just a particular configuration of particular parts of physical computer memory. Certain configurations of open/closed transistors, or high/low current are understood as representing certain binary digits 1 or 0 which are interpreted as numbers.

[28] Charles Petzold does an excellent job of explaining computer architecture in his book *Code* (Petzold 1999).

[29] Annoyingly, the "+" operator is often overloaded to do string concatenation, instead of adding

I think the same sort of thing is going on in the relationship between the decision theory frameworks. Everything can be thought of, at base, as a proposition. But some kinds of propositions we want to think of as events, and their logical structure and interrelationships are important. For other propositions, it is the preference structure which is important to us.

Often, the abstraction of treating different kinds of propositions as different is a legitimate move: things are conceptually simpler. Savage sets things up in this simple way. But it is a "leaky abstraction" as all abstractions are.[30] What this means is that in the edge cases, in the unusual circumstances, the abstraction breaks down. And when this happens, all sorts of strange behaviour can emerge.

As long as you are aware of the problems and do your best to avoid them, then you can safely use the abstracted theory. Thus I adopt Savage's strict distinctions between acts, states and outcomes.

Bradley (2007) explains exactly what needs to be done to a "Jeffrey-like" framework to make it look "Savage-like". This amounts to treating acts as conjunctions of particular kinds of conditionals in an enriched space. These are not truth-functional conditionals, but there is some disagreement about what sort of conditionals they are. Joyce (1999) offers one approach, Bradley (2007) another.

So, acts will now be thought of as functions from states to outcomes. But we should keep in mind that they are "really" conjunctions of conditionals of a certain kind.

2.5. Conclusion

We have now set up the formalism that will see us through the rest of the dissertation. It is abstract, and idealised in many ways, as we have seen. However, the hope is that some insight can be drawn from this system nevertheless.

the underlying numbers together.

[30]The term is from Spolsky (2002)

3. Arguments for probabilism

> The long run is a misleading guide to current affairs. In the long run we are all dead. Economists set themselves too easy, too useless a task if in tempestuous seasons they can only tell us that when the storm is past the ocean is flat again.
>
> *(John Maynard Keynes)*

There were quite a few preliminary details to set out in the previous chapter, so it may be useful to take a moment to explain how this chapter fits into the overall project. The ultimate aim is to develop a better account of belief and decision under severe uncertainty. Before I do this, it will be helpful to see what the standard approach to these issues is. I take probabilism to be the standard approach and so this chapter is devoted to discussing and criticising the arguments for probabilism.[1] The next chapter will contain further discussion of probabilism, and will present some further problems for the standard view.

In section 3.1 I discuss one of the most influential arguments for probabilism: the Dutch book theorem. Next, in section 3.2 I tackle the other main type of argument: representation theorems. Having dealt with these two well discussed strands of justification for probabilism, I move on to some less well known views. A recent kind of argument that relies on measuring inaccuracy of belief is summarised in section 3.3. Finally, in section 3.4 I discuss some other attempts to justify probabilism.

There are a couple of things to explain here. First "coherent" often means conforming to the rules of probability. So incoherent credences are those that

[1] Recall that probabilism is the view that takes probability theory to be the best model of belief under uncertainty.

contravene one or other of the axioms of probability. I don't like using "coherent" to mean "conforming to probability theory", since the suggestion is then that non-probabilists are incoherent. This is a value judgement I do not endorse. So perhaps I should call probabilists "probabilistically coherent" or just "p-coherent". One could then talk of other kinds of coherence: "DS-coherent", meaning conforming to the axioms for Dempster-Shafer beliefs.

Sometimes probabilism is taken to include the various alternative approaches I outlined above. My use of the term "probabilism" is somewhat more restricted: I use it to mean coherence with probability theory as outlined in Definition 2.3.3. I do however endorse the more general claim that degrees of belief can and should be modelled formally. In fact, I even endorse the claim that probability theory has a privileged role to play in our understanding of degrees of belief. I return to this point at the end of the chapter.

3.1. Dutch book argument

An awful lot of ink has been spilled over the Dutch book argument (DBA).[2] People have criticised its prudential character, that it relies on pragmatic concerns about preferring higher utility outcomes. People have suggested that betting behaviour is not a good proxy for belief (e.g. Bradley and Leitgeb 2006). Any number of other issues have been raised with it. See Hájek (2008) for an overview. I am sidestepping much of the current and past debate over the status of the Dutch book argument. I want to explore in detail what one needs to assume in order to prove the theorem, and ask when are those assumptions justified.

Degrees of belief have a role in decision making, and betting is a paradigm case of that. So there should be *some* connection between betting behaviour and belief. The connection may not be as tight as proponents of the DBA would have us believe, but there is still something to be learned about belief by studying the constraints on reasonable betting.

Standardly the theorem is traced back to the work of Bruno De Finetti and F.P. Ramsey in the 20s and 30s.[3] My aim here is not historical, so I do not want to go back to those original authors and see whether modern treatments are faithful to those pioneers' works. What I will do is take one specific modern treatment

[2] Much of this section draws on Bradley (2012)

[3] I am here making a claim about where the argument is generally traced to: not one about who it *should* be traced to.

of the theorem and go through it in detail. I will also mention another different approach to the theorem.

As a first approximation the Dutch book argument says that if your credences are "non-probabilistic" then you are open to a Dutch book. Non-probabilistic means just that your credences don't conform to the axioms of probability. A Dutch book is a set of bets that guarantee a sure loss to whoever takes them. But strictly speaking the Dutch book argument does not say this: it talks in terms of betting behaviour and implicitly assumes that betting behaviour lines up with belief. So, let's restate the DBA in terms of betting behaviour. It says if your betting odds are non-probabilistic then you are open to a Dutch book and you will lose money. So, if your degrees of belief are the sole determinant of your betting odds, then your degrees of belief had better be probabilistic.

But this can't be the whole story, since actual bookmakers *do* offer odds that are not probabilistic! If this naïve Dutch book argument were true, then bookies offering nonprobabilistic odds would have been exploited to bankruptcy long ago. So there must be more to the story than that. To give a fairer appraisal of DBA, I need to explore in detail the theorem that gives the argument its strength.

This will be a rather long and involved discussion of what is, ultimately, a rather simple theorem. The aim is to carefully assess exactly what restrictions on betting behaviour are required to secure the theorem. Along the way, we will see which parts of the theorem can be proven with restricted sets of constraints. Since I will find only some of the conditions to be rationally compelling, it is worthwhile seeing what can be done without the contentious axioms.

3.1.1. A formal framework for gambling

The Dutch book argument talks in terms of betting behaviour, so we need a formal method for talking about bets. I borrow much of this formalism from Halpern (2003). Frank Döring (2000) has a slightly more general framework for describing bets, but when he moves to discussions of Dutch books, he places restrictions on the allowable bets so as to make the two frameworks much the same.

For our purposes, a bet is an ordered pair of an event in SL and a "betting quotient". For a bet of stake s and potential winnings w, the betting quotient is $\alpha = \frac{s}{w}$. The higher the α the more likely you think the event in question is. The greater the proportion of the winnings you are willing to risk on a bet, the more likely you think the event is. So bets will be ordered pairs of the form (X, α) where $X \in SL$ and $\alpha \in \mathbb{R}$. What is relevant about a betting scenario is the betting quotient and the event in question.

The argument can be made that the absolute size of the stake and the potential winnings relative to your current wealth are also relevant in gambling. For example, in poker, a good player will act differently depending on whether she is chip-leader or short-stacked. The way Joyce (1999) sets up his decision theory builds this element in. Joyce makes you sensitive to overall levels of wealth, rather than changes in level of wealth. While this is better in actual cases, in this rarefied setting we are assuming that utility is linear with money and that the punter is risk neutral. This is for the sake of simplicity, rather than for any principled reason. Concerns about the diminishing marginal utility of money, and your attitude to risk are orthogonal to the concerns of the current project. Severe uncertainty is hard enough on its own, so I will help myself to these sorts of simplifications throughout. I am not making any claims that these simplifications are always permissible, simply that the complications they lead to are orthogonal to the concerns of the current project.

A bet (X, α) pays out $\mathbf{v}(X)$. That is, if X is true the bet pays out 1, and it pays out 0 if X is false. The bet costs α, that is to say the *stake* is α because we have set the winnings $w = 1$. The bet $(\neg X, 1-\alpha)$ is called the *complementary bet* to (X, α). Think of the complementary bet $(\neg X, 1-\alpha)$ as "selling" the bet (X, α). Whenever you take a bet (X, α), the bookie is effectively taking on the complementary bet $(\neg X, 1-\alpha)$. Table 3.1 illustrates the "mirror image" quality that the payoffs of complementary bets have.

	$\mathbf{v}(X) = 1$	$\mathbf{v}(X) = 0$
(X, α)	$1-\alpha$	$-\alpha$
$(\neg X, 1-\alpha)$	$-(1-\alpha)$	α

Table 3.1.: Payoffs for a bet and its complement

What I am interested in here is how betting behaviour might serve to force degrees of belief to be probabilistic. In this respect, the important parts of the bet are the event bet on and something that is supposed to correlate with the agent's strength of the belief in that event. The betting quotient serves this purpose.

Think of the bets like this: when you buy a bet (X, α) you pay α and get a ticket that says X. If it turns out that X is indeed the case – that is it turns out that $\mathbf{v}(X) = 1$ – then you return to the bookie and redeem your ticket. That is, you hand in your ticket to the bookie and he gives you £1. So your net gain is $1 - \alpha$. If X fails to be the case – if $\mathbf{v}(X) = 0$ – then the ticket is worthless and you have lost your α.

There are two possibilities for the "neutral bet": $(\top, 1)$ or $(\bot, 0)$. That is, the bet that costs 1 and guarantees a payoff of 1 or the free bet with no prospect of winning anything. Preferring a bet to one of these suggests that you would accept that bet: the bet is "better than nothing". I return to the issue of preference versus acceptance later. These neutral bets bets are complementary to each other.

A set of bets $B = \{(X_1, \alpha_1), \dots (X_k, \alpha_k)\}$ costs $\kappa(B) = \sum \alpha_i$ and pays out $\pi_{\mathbf{v}}(B) = \sum \mathbf{v}(X_i)$. That is, you get 1 for every event that you bet on that is true. So the value of a set of bets B is $\tau_{\mathbf{v}}(B) = \pi_{\mathbf{v}}(B) - \kappa(B)$. The value of the bet is how much it pays out, minus what the bet cost. The complementary set of bets[4] B^C is just the set of bets complementary to those in B: $\{(\neg X_1, 1 - \alpha_1) \dots (\neg X_k, 1 - \alpha_k)\}$

Theorem 3.1.1 *For all* $\mathbf{v} \in \mathbf{V}$ *and* B *we have* $\tau_{\mathbf{v}}(B^C) = -\tau_{\mathbf{v}}(B)$

Proof Assume that B consists of n bets: $\{(X_1, \alpha_1), (X_2, \alpha_2), \dots, (X_n, \alpha_n)\}$. The cost of the complementary bet is:

$$\begin{aligned}
\kappa(B^C) &= (1 - \alpha_1) + (1 - \alpha_2) + \dots + (1 - \alpha_n) \\
&= n - \sum \alpha_i \\
&= n - \kappa(B)
\end{aligned}$$

As for the payout: for every bet that B wins, B^C does not, and for every bet that B does not win, B^C does. There is a total of n up for grabs and exactly one of B or B^C wins on each of the n events. More concisely, from the properties of $\mathbf{v} \in \mathbf{V}$, we know that $\mathbf{v}(\neg X_i) = 1 - \mathbf{v}(X_i)$. So:

$$\begin{aligned}
\pi_{\mathbf{v}}(B^C) &= \sum \mathbf{v}(\neg X_i) \\
&= \sum (1 - \mathbf{v}(X_i)) \\
&= n - \sum (\mathbf{v}(X_i)) \\
&= n - \pi_{\mathbf{v}}(B)
\end{aligned}$$

Putting these two results together it follows that a set of bets wins exactly as

[4] This is an abuse of notation. B^C is not the set-theoretic complement of B at all, but something rather more subtle.

much as the complementary set of bets loses:

$$\begin{aligned}\tau_{\mathbf{v}}(B^C) &= (n - \pi_{\mathbf{v}}(B)) - (n - \kappa(B))\\ &= -\pi_{\mathbf{v}}(B) + \kappa(B)\\ &= -\tau_{\mathbf{v}}(B)\end{aligned}$$

■

This shows that the neat complementarity of the payouts of complementary bets illustrated in Table 3.1 holds for *sets* of complementary bets as well.

(X, α) is the same as a bet with odds of "$1 - \alpha : \alpha$". This is for English-style odds. That is, "odds-to" rather than "odds-for".[5] Apparently in the U.S. odds look different and the corresponding bet would have odds of "$1 : \alpha$".

Odds can be defined directly in terms of bets; as a ratio of the profit to the stake. Profit is the win minus the stake. So for the bet (X, α), a stake of α gives a profit of $1 - \alpha$ if X obtains. To tie this in with the above discussion of α as a betting quotient, recall that our potential winnings are fixed at 1. So α plays the role of both the betting quotient and the stake. $\alpha = \frac{s}{w} = \frac{\alpha}{1}$. There is no loss of generality here, since if you wanted to play a "high stakes" gamble with large potential winnings, say 100, you can simply buy one hundred copies of the bet. You could just buy 100 copies of (X, α). One could make the size of the stake an extra parameter (as Paris (1994, 2005 [2001]) does) and then add a constraint to the effect that preferences among bets are insensitive to the size of the stakes. That is, a bet would be a triple (X, α, s) that costs $s\alpha$ and pays out $\alpha\,\mathbf{v}(X)$. The constraint would then be that if $(X, \alpha, s) \succ (X', \alpha', s)$ then the same preference holds for all s'. A more formal rendering of this sort of "stake independence" appears as Savage's P4 in a later section.

Recall the standard caveats from section 2.4.1. We want any preference among bets to indicate a difference in the punter's degrees of belief in those events. That is, we want to say that if a punter prefers bet A to bet B when A and B pay out the same, then that reflects the punter's belief that the event in A is more likely than the event in B.

3.1.2. Rationality constraints on betting preference

I am now ready to set out the proof of the Dutch book theorem. This section lays out the preliminaries, the constraints on betting preference. The next section contains the proof proper. First I will make clear exactly what conditions have

[5] The odds-to of a fair coin landing heads should be "1 to 1". Odds-for would be "2 for 1".

to hold for the argument to follow. That is, I show what conditions have to be satisfied in order for you to be constrained by the DBA. These conditions come from Halpern (2003). This treatment is nonstandard. I chose Halpern's treatment because it is the one that makes clearest the assumptions one makes in proving the theorem. They are also interestingly weaker than might be expected. I will later discuss how it relates to other more classical outlines.

The preference relation we are characterising is one that holds between sets of bets, but this straightforwardly determines a relation between the bets themselves by its action on singleton sets, so I use the "$\succeq$" symbol for relations among individual bets as well.

Halpern's first constraint is the following:

> Bet Dominance *For sets of bets B_1 and B_2: If $\tau_{\mathbf{v}}(B_1) \geqslant \tau_{\mathbf{v}}(B_2)$ for all $\mathbf{v} \in \mathbf{V}$ then $B_1 \succeq B_2$. And if $\tau_{\mathbf{v}}(B_1) > \tau_{\mathbf{v}}(B_2)$ for some $\mathbf{v}$ then $B_1 \succ B_2$.*

This is a dominance principle. I call it Bet Dominance to distinguish it from a principle called Dominance that comes up in the gradational accuracy argument (section 3.3). What this says is that if B_1 is guaranteed to give at least as much money as B_2 in every situation, then you should (weakly) prefer B_1 to B_2. And if B_1 sometimes gets you strictly more money, you should strictly prefer it. This condition relates preference among bets to payoffs of bets. It is this condition that gives the theorem its pragmatic character. Important steps in the proof work by showing that if you are in violation of one of the axioms of probability, then it can be shown that you prefer a sure loss to a sure gain, thus contradicting Bet Dominance. Note that this condition does not entail that you are rationally obliged to prefer a bet with higher expected value: you are only obliged to prefer a bet with higher *guaranteed* value. This is one way Halpern's conditions are weaker than might be expected. Indeed, it would be strange if Halpern's argument relied on expectation: this is supposed to be a behaviourist argument, so it should not appeal to the inner mental life of the agents and what they expect.

Halpern's second condition is simply that betting preferences should be transitive.

> Transitivity *For sets of bets B_1, B_2, B_3: if $B_1 \succeq B_2$ and $B_2 \succeq B_3$ then $B_1 \succeq B_3$*

This second condition gives some structure to the set of bets. Transitivity and the fact that the weak preference relation is reflexive[6] means that the relation gives the set of bets the structure of a partial preorder.

[6] Reflexivity follows from Bet Dominance.

The third condition is that complementary bets are comparable.

Complementarity *For all X either $(X,\alpha) \succeq (\neg X, 1-\alpha)$ or $(\neg X, 1-\alpha) \succeq (X,\alpha)$*

This condition relates complementary events. In other words it relates betting behaviour regarding an event to behaviour regarding the complementary event. Or it connects your attitude to buying bets to your attitude to selling bets. This gives further structure to the betting preferences. Since we are using betting behaviour as a way of discussing uncertainty, this condition relates uncertainty in an event to uncertainty in its complement. Note this is an inclusive "or": both disjuncts of (Complementarity) can hold.[7]

Note that this condition is much weaker than demanding that the preference relation be complete, or total. It is only between *complementary* bets that the relation must hold one way or the other. In other words, it can still be acceptable, rational, that neither $B_1 \succeq B_2$ nor $B_2 \succeq B_1$, as long as the bets are not complementary. This is the condition I will end up rejecting. I shall defer my criticisms of it until the end of this section.

The final condition states that if you prefer some bets individually, you prefer them as a set. Or, as Halpern puts it "preferences are determined point-wise". (Halpern 2003, p. 22)

Package *If $(X_i,\alpha_i) \succeq (Y_i,\beta_i)$ for each i then:*

$$\{(X_1,\alpha_1),\dots(X_k,\alpha_k)\} \succeq \{(Y_1,\beta_1),\dots(Y_k,\beta_k)\}$$

This relates single bets to sets of bets. It is sometimes called the "package principle" and has been criticised (see e.g. Schick 1986). I think it's fairly acceptable: I don't feel the force of the objections against it. I will however show what it is possible to prove in the absence of this condition.

This constraint is fairly weak. Note the things it does *not* say: it does not say that if $B_1 \succeq B_2$ then for each $(X_1,\alpha_1) \in B_1, (X_2,\alpha_2) \in B_2$ that $(X_1,\alpha_1) \succeq (X_2,\alpha_2)$. Nor does it guarantee in general that if $B_1 \succeq B_2$ and $B_3 \succeq B_4$ then $B_1 \cup B_3 \succeq B_2 \cup B_4$.

These constraints impose a certain structure on an agent's preference among bets. This concludes the discussion of the constraints on rational betting preference. It

[7] However, it can be shown that both disjuncts of (Complementarity) are true for at most one value of α for any X.

is claimed that failing to have preferences structured in this way is irrational. The intuitive appeal of most of the conditions should be apparent.

Real bookies violate (Complementarity). They do not offer "symmetric" odds on and against an event.[8] But of course, real bookies' odds do not reflect their beliefs in the way agent's preferences in Dutch book situations should. Bookies are out to make money. The move away from symmetry allows bookies to make money whichever way the outcome goes. They are, in effect, collectively Dutch booking their punters. This assumption of Dutch book arguments is not often made explicit, but it is important: agents in the DBA are not out to make money, nor are they trying to lose money; they simply bet according to their beliefs in order to break even. That is, these agents are setting their betting quotients in such a way as to reflect what they consider fair prices. The agents are like a cooperative insurance agency: they want to set their premiums in such a way as to cover the losses they expect, but they also want to set their premiums as low as possible.[9] That's why their betting behaviour is supposed to tell us about their beliefs. That's why (Complementarity) is supposed to be justified. I shall have more to say on this point later.

3.1.3. Consequences of betting preference conditions

The following two definitions will be absolutely vital to the proof.

Definition 3.1.1 *Your* limiting betting quotients *are defined:*

- $\alpha_X = \sup\{\alpha : (X, \alpha) \succeq (\neg X, 1 - \alpha)\}$
- $\beta_X = \inf\{\beta : (\neg X, 1 - \beta) \succeq (X, \beta)\}$

That is, α_X is the largest value you're willing to buy the bet at, and β_X is the smallest value at which you'll sell the bet. The Dutch book theorem, in its usual form amounts to arguing that $\mathbf{pr}(X) = \alpha_X$ is a probability measure. That is, we need to show that $\alpha_\top = 1$; $\alpha_X \geqslant 0$ for every X; and that $\alpha_{X\vee Y} = \alpha_X + \alpha_Y$ for incompatible X, Y.[10]

I don't do this directly. I try to do as much of the proof as possible with as few of the conditions as possible. It turns out that you can do a lot without

[8]Betting quotients that satisfy this condition are symmetric in the sense that if the odds on X are $1 - \alpha : \alpha$ then the odds against X are $\alpha : 1 - \alpha$.
[9]This is an analogy that is due to Lenny Smith.
[10]This is using the "incompatible propositions" definition of probability.

(Complementarity), and then use it just at the end to show $\alpha_X = \beta_X$. And then the last details fall into place.

I set out the proof as follows. First I show that only the first two conditions – (Bet Dominance) and (Transitivity) – are required to give $\underline{\mathbf{pr}}(X) = \alpha_X$ and $\overline{\mathbf{pr}}(X) = \beta_X$ the structure of a Choquet capacity. Then I show that with (Package) as well I can demonstrate that $\underline{\mathbf{pr}}(X) = \alpha_X$ is a Dempster-Shafer belief function and that $\overline{\mathbf{pr}}(X) = \beta_X$ is a Dempster-Shafer plausibility function. Finally, I introduce (Complementarity) and demonstrate that the full Dutch book theorem now holds.

The proof is not "efficient" in that the theorem can be proved much quicker if all the conditions are available from the start. However, I think that the intermediate results that I derive on the way are interesting in their own rights. And since I don't endorse all the constraints on betting preference, those intermediate results will be all that will be required for rationality, in my view.

What the Dutch book theorem actually proves

The Dutch book theorem shows us that limiting betting quotients must have a certain structure if your betting preferences satisfy the above restrictions. Typically the Dutch book theorem is taken to tell us something about *belief*, not just about betting quotients. To make this move, we need one further assumption. We need something that makes the link between betting quotients and belief tight enough that structure imposed on betting quotients allows us to infer the same structure imposed on belief. One way to do this is to *interpret* degrees of beliefs as your limiting willingness to bet. This is the line that Williamson (2010) takes. I think this is too strong a connection between belief and betting behaviour. Obviously revealed betting preferences have some strong connection to belief, but I don't think that betting preference exhausts the functional role that belief plays. I discuss in more detail what I understand belief to be when I discuss my preferred model in the next chapter (section 4.2).

Despite this, it does seem that there should be a strong link between belief and betting behaviour. So we shall assume that you are willing to bet in line with your beliefs. To secure the conclusions we want, we need to be able to assume that betting behaviour reveals the structure of your beliefs.

One might think that there are distinctions in belief state that can't be elicited by differences in betting behaviour. Even so, the Dutch book theorem tells us something about your betting quotients. This tells us *something* about your beliefs, but we mustn't make the mistake of thinking that betting quotients *are* beliefs:

they are consequences of beliefs, only. So the following sections are going to put the Dutch book theorem in terms of betting quotients. It will then be up to us to decide what the consequences of this are for belief. How we do this depends on what we think the relationship is.

Betting quotients are Choquet capacities

To show that betting quotients are Choquet capacities[11] we need to show that they are normalised, bounded and monotonic. Normalisation first.

Lemma 3.1.1 $\alpha_\bot = 0 \qquad \alpha_\top = 1 \qquad \beta_\bot = 0 \qquad \beta_\top = 1$

Proof Note that $\tau_{\mathbf{v}}(\bot, \alpha) = -\alpha$ for all $\mathbf{v}$. Likewise $\tau_{\mathbf{v}}(\top, 1-\alpha) = \alpha$ for all $\mathbf{v}$. For $\alpha < 0$, this means that $(\bot, \alpha) \succ (\top, 1-\alpha)$ by Bet Dominance. Thus $\alpha_\bot \geqslant 0$. For $\alpha < 0$, we have $(\top, 1-\alpha) \succ (\bot, \alpha)$. So $(\bot, \alpha) \not\succeq (\top, 1-\alpha)$. So $\alpha_\bot \leqslant 0$. Thus $\alpha_\bot = 0$. Proofs of the other three results follow in exactly the same way. ∎

So our betting quotients are normalised.

We move on to monotonicity now. First we have a couple of quick lemmas regarding preference for smaller bets and for bigger events. Note that many of these lemmas also have a "strict" version, but I leave that implicit, since I never actually need to use the strict versions.

Lemma 3.1.2 *For all events, X:*

- *If $\alpha \leqslant \alpha'$ then $(X, \alpha) \succeq (X, \alpha')$*
- *If $\beta \geqslant \beta'$ then $(\neg X, 1-\beta) \succeq (\neg X, 1-\beta')$*

Proof This follows immediately from (Bet Dominance). ∎

That is, you prefer the bet that costs you less if the bets win/lose in the same circumstances. And similarly, if you're selling the bet, you prefer to sell it for more money.

Lemma 3.1.3 *If $X \models Y$ then for all α:*

- $(Y, \alpha) \succeq (X, \alpha)$
- $(\neg X, 1-\alpha) \succeq (\neg Y, 1-\alpha)$

[11] Choquet capacities were defined above in Definition 2.3.5, p. 34.

Proof The first half follows directly from (Bet Dominance). The second half follows from the first half and the observation that if $X \models Y$ then $\neg Y \models \neg X$. ■

This too should be intuitive. If you keep the stake fixed, you would prefer a bet that wins in more circumstances. And if you are selling the bet it is the other way around: you prefer to sell a bet on a smaller event.

Lemma 3.1.4 *For all events, X, the following two biconditionals hold:*

- $(X,\alpha) \succeq (\neg X, 1-\alpha)$ *if and only if* $\alpha \leqslant \alpha_X$
- $(\neg X, 1-\beta) \succeq (X,\beta)$ *if and only if* $\beta \geqslant \beta_X$

Proof We shall do the "if" part first; we assume $\alpha \leqslant \alpha_X$.

	$(X,\alpha) \succeq (X,\alpha_X)$	by Lemma 3.1.2
	$(X,\alpha_X) \succeq (\neg X, 1-\alpha_X)$	by definition
	$(\neg X, 1-\alpha_X) \succeq (\neg X, 1-\alpha)$	by Lemma 3.1.2
So:	$(X,\alpha) \succeq (\neg X, 1-\alpha)$	by (Transitivity)

Now the "only if" part. (X,α) is preferred to its complement. It follows that $\alpha \leqslant \alpha_X$ because α_X is *defined* as the largest value for which that bet is preferred to its complement. The proof of the second result is similar. ■

The above lemmas together allow us to prove the following:

Theorem 3.1.2 *If* $X \models Y$ *then:*

$$\alpha_X \leqslant \alpha_Y \qquad \textit{and} \qquad \beta_X \leqslant \beta_Y$$

Proof Assume $X \models Y$. Then:

	$(Y,\alpha_X) \succeq (X,\alpha_X)$	by Lemma 3.1.3
	$(X,\alpha_X) \succeq (\neg X, 1-\alpha_X)$	by definition
	$(\neg X, 1-\alpha_X) \succeq (\neg Y, 1-\alpha_X)$	by Lemma 3.1.3
	$(Y,\alpha_X) \succeq (\neg Y, 1-\alpha_X)$	by Transitivity
So:	$\alpha_X \leqslant \alpha_Y$	by Lemma 3.1.4

The proof of the other result proceeds in the same way. ■

Since $\bot \models X$ and $X \models \top$ for all X, monotonicity entails boundedness. We have proved that $\underline{\mathbf{pr}}(X) = \alpha_X$ has the structure of a Choquet capacity, as does $\overline{\mathbf{pr}}(X) = \beta_X$.

We can in fact prove more than just that $\underline{\mathbf{pr}}(X) = \alpha_X$ and $\overline{\mathbf{pr}}(X) = \beta_X$ are capacities. We can show that they are capacities intimately related to each other. It is hardly surprising that values at which you buy a bet and values at which you sell those bets are connected.

Theorem 3.1.3 *For all X; $\alpha_X = 1 - \beta_{\neg X}$*

Proof Note that $\neg\neg X = X$ and $1 - (1 - \beta_{\neg X}) = \beta_{\neg X}$, and thus that $(X, 1 - \beta_{\neg X}) \succeq (\neg X, \beta_{\neg X})$. So by Lemma 3.1.4 we know that $1 - \beta_{\neg X} \leqslant \alpha_X$.

$1 - (1 - \alpha_X) = \alpha_X$. So from the definition of α_X it follows that $(\neg\neg X, 1 - (1 - \alpha_X)) \succeq (\neg X, 1 - \alpha_X)$. Therefore, by Lemma 3.1.4, $1 - \alpha_X \geqslant \beta_{\neg X}$. Thus, $1 - \beta_{\neg X} \geqslant \alpha_X$.

Therefore $\alpha_X = 1 - \beta_{\neg X}$ ■

In fact, if we allowed our "αs" to range over some only partially ordered set of goods with unique maximal and minimal elements, rather than real numbers, then one can prove that (Bet Dominance) and (Transitivity) are enough to show that the belief function so defined is Halpern's "plausibility measure".[12] See Halpern (2003, p. 50–54) or Chu and Halpern (2008) for details.

Betting quotients are Dempster-Shafer belief measures

All of the above results rely only on (Bet Dominance) and (Transitivity). From now on, the results will also require (Package).[13]

Theorem 3.1.4 *For all X; $\alpha_X \leqslant \beta_X$*

Proof The set of bets $\{(X, \alpha_X), (\neg X, 1 - \beta_X)\}$ is preferred to its complement by definition of α_X and β_X and (Package). The payout of this set of bets is always 1. The cost of the bets together is $\alpha_X + 1 - \beta_X$. So the value of this set of bets is always $\beta_X - \alpha_X$. If it were the case that $\alpha_X > \beta_X$, this net gain would be negative, and the complementary set of bets would have positive net gain, by Theorem 3.1.1. This contradicts (Bet Dominance), so the conclusion is that $\alpha_X \leqslant \beta_X$ ■

This is a proof strategy that we shall use a lot: we make some assumptions about some betting quotients. We find some set of bets (often a set of bets with one bet on each element of a partition of SL) and we show that this set is preferred to its complement but that it has a negative payout in every world (that is, for every $\mathbf{v}$).

[12] The definitions of $\tau_{\mathbf{v}}$ and friends need tweaking, but all the modifications are straightforward.

[13] In fact, the next few results require only (Bet Dominance) and (Package). Until we get to Theorem 3.1.5, which also requires (Transitivity), no interesting structure is imposed.

So its complement has a positive payout in every world by Theorem 3.1.1. This violates (Bet Dominance), so whatever assumption we made about the αs must be wrong. Often I will show that the set has a negative payout and jump straight to the conclusion that the assumption is wrong.

Our next aim is to show that α_X is *superadditive* and that β_X is *subadditive*. That is, we want to demonstrate that the following two inequalities hold for disjoint X and Y: $\alpha_{X\vee Y} \geqslant \alpha_X + \alpha_Y$ and $\beta_{X\vee Y} \leqslant \beta_X + \beta_Y$. Before we do that, we will need some more lemmas.

Lemma 3.1.5 *If exactly one of the X_i are true then:*

$$\sum \alpha_{X_i} \leqslant 1 \qquad \textit{and} \qquad \sum \beta_{X_i} \geqslant 1$$

Proof Each (X_i, α_{X_i}) is preferred to its complement by definition. So by (Package) they are preferred as a set. This set of bets always wins 1, no matter what $\mathbf{v}$ is used, since exactly one X_i is true. If $\sum \alpha_{X_i} > 1$ then the net gain would be negative, contradicting (Bet Dominance). The proof of the second half is similar. ■

Note that with (Complementarity) and Lemma 3.1.5 it is already possible to prove that $\mathbf{pr}(X) = \alpha_X$ conforms to the partition definition of probability.

Note that exactly one of $\{X \vee Y, \neg(X \vee Y)\}$ is true and therefore, by the above lemma, that $1 \geqslant \alpha_{X\vee Y} + \alpha_{\neg(X\vee Y)}$.

Two final inequalities will allow us to prove the result we want.

Lemma 3.1.6

$$\alpha_{X\vee Y} + \alpha_{X\wedge Y} \leqslant \beta_X + \beta_Y \qquad \textit{and} \qquad \beta_{X\vee Y} + \beta_{X\wedge Y} \geqslant \alpha_X + \alpha_Y$$

Proof The following set of bets is preferred to its complement by definition (plus (Package)): $\{(X \vee Y, \alpha_{X\vee Y}), (X \wedge Y, \alpha_{X\wedge Y}), (\neg X, 1-\beta_X), (\neg Y, 1-\beta_Y)\}$. This set of bets wins 2 everywhere. Its cost is $(1-\beta_X) + (1-\beta_Y) + \alpha_{X\wedge Y} + \alpha_{X\vee Y}$. Thus the set of bets wins $\beta_X + \beta_Y - (\alpha_{X\vee Y} + \alpha_{X\wedge Y})$ everywhere. So in order to avoid this being negative, $\alpha_{X\vee Y} + \alpha_{X\wedge Y} \leqslant \beta_X + \beta_Y$. The other half of the proof is similar. ■

We now have everything we need to prove the following theorem:

Theorem 3.1.5 *For all incompatible X, Y:*

- $\alpha_{X\vee Y} \geqslant \alpha_X + \alpha_Y$
- $\beta_{X\vee Y} \leqslant \beta_X + \beta_Y$

Proof Consider the bet $B = (X \vee Y, \alpha_X + \alpha_Y)$. This has exactly the same payout as $C = \{(X, \alpha_X), (Y, \alpha_Y)\}$. Because X and Y are incompatible, C never wins 2. So by (Bet Dominance), $B \succeq C$ (and indeed, $C \succeq B$). $C \succeq C^C$ by definition and (Package). Now by (Bet Dominance) again, $C^C \succeq B^C$. So by (Transitivity) $B \succeq B^C$. $\alpha_{X \vee Y}$ is defined as the maximum value for which a bet on $X \vee Y$ is preferred to its complement. This value is at least $\alpha_X + \alpha_Y$, since $B \succeq B^C$. So $\alpha_{X \vee Y} \geqslant \alpha_X + \alpha_Y$.

Much the same proof works for the corresponding result involving $\beta_{X \vee Y}$. ■

Lemma 3.1.6 and Theorem 3.1.5 don't quite get us what we'd really like to prove. To prove a more general result directly – namely that $\alpha_{X \vee Y} + \alpha_{X \wedge Y} \geqslant \alpha_X + \alpha_Y$ – we need another constraint on preference among sets of bets. I have proved these slightly unsatisfactory theorems because they, together with Complementarity, are sufficient to prove what we want without recourse to the extra principle. The required principle is the following:

Distribution Invariance *For set of bets* $B = \{(X_1, \alpha_1), (X_2, \alpha_2), \ldots, (X_n, \alpha_n)\}$, *if* $B \succeq B^C$ *then* $\kappa(B) \leqslant \sum \alpha_{X_i}$.

This principle requires that preference among sets of bets is determined *only* by the total cost of the set and not how that cost is distributed among the individual bets. The converse of this principle is a consequence of Bet Dominance.

Lemma 3.1.7 *For all* X, Y

$$\alpha_X + \alpha_Y \leqslant \alpha_{X \vee Y} + \alpha_{X \wedge Y}$$

Proof Let $B = \{(X \vee Y, \alpha_X), (X \wedge Y, \alpha_Y\}$ and let $C = \{(X, \alpha_X), (Y, \alpha_Y)\}$. $\tau_{\mathbf{v}}(B) = \tau_{\mathbf{v}}(C)$ for all $\mathbf{v}$. So by Bet Dominance $B \sim C$. Also, $\tau_{\mathbf{v}}(B^C) = \tau_{\mathbf{v}}(C^C)$ so $B^C \sim C^C$. $C \succeq C^C$ by definition and Package. Putting these facts together with Transitivity we have that $B \succeq B^C$. So by Distribution Invariance we have that $\alpha_X + \alpha_Y \leqslant \alpha_{X \vee Y} + \alpha_{X \wedge Y}$. ■

The same proof can be used to show that the αs are n-monotonic for any n. If we define $\mathbf{bel}(X) = \alpha_X$ and $\mathbf{plaus}(X) = \beta_X$ these functions are Dempster-Shafer belief and plausibility measures, respectively.

When we introduced Dempster-Shafer belief functions,[14] we also introduced the connected notion of a *mass function*. We can construct the mass function out

[14]In section 2.3.2 on p. 2.3.8.

of our betting quotients. First, let's define $m(X)$ as follows:

$$m(X) = \alpha_X - \sum_{Y \not\dashv\models X} m(Y) \tag{3.1}$$

By "$Y \not\dashv\models X$" I mean Y entails X but X does not entail Y. That is, $[Y] \subsetneq [X]$. Strictly speaking this should be an inductive definition on the size of X, but it's fairly clear how it works. This will turn out to be a mass function as defined in Definition 2.3.9 on p. 36. Note that this sum is "ranging over proper subsets of $[X]$". $m(X)$ picks up all the mass not already attributed to subsets of X. So for example, the mass of an atom is just equal to its maximal betting quotient. That is $m(t) = \alpha_t$. From this assumption and Lemma 3.1.3 it follows that $m(X) \geqslant 0$ for all X; because $\bot \models X$, for all X. It just remains for us to show that the sum of the masses of all events is equal to 1. This follows from the fact that the atoms partition the event space, and thus $\sum \alpha_{t_i} \leqslant 1$ by Lemma 3.1.5.

We now show that the DS belief function defined by this mass function does indeed have the right relationship to all all the betting quotients.

Theorem 3.1.6 *For all X:* $\mathbf{bel}(X) = \alpha_X$.

Proof By definition:

$$\begin{aligned}
\mathbf{bel}(X) &= \sum_{Y \models X} m(Y) \\
&= m(X) + \sum_{Y \not\dashv\models X} m(Y) \\
&= \alpha_X - \sum_{Y \not\dashv\models X} m(Y) + \sum_{Y \not\dashv\models X} m(Y) \\
&= \alpha_X
\end{aligned}$$

■

We can also prove, in a similar way, that $\mathbf{plaus}(X) = \beta_X$ is a D-S plausibility function, though I won't go through that here. So **bel** defined this way is infinite monotone. We have proved as much as we can without (Complementarity).

Betting quotients are probabilities

Finally, we add (Complementarity) to our arsenal of rationality constraints and now we can prove the full Dutch book theorem.

Lemma 3.1.8 *For all X, $\alpha_X = \beta_X$.*

Proof For all $\beta > \alpha_X$, it is not the case that $(X,\beta) \succeq (\neg X, 1-\beta)$. So by (Complementarity), $(\neg X, 1-\beta) \succeq (X,\beta)$. So $\alpha_X = \inf\{\beta : (\neg X, 1-\beta) \succeq (X,\beta)\} = \beta_X$ by Definition 3.1.1. ■

So our superadditive α_X and our subadditive β_X are forced to be the same. This is enough to derive the final result.

Theorem 3.1.7 $\mathbf{pr}(X) = \alpha_X$ *is a probability measure.*

Proof By Lemma 3.1.6, and the above result, we know that $\alpha_{X \vee Y} = \alpha_X + \alpha_Y$ for incompatible X, Y. So $\mathbf{pr}(X) = \alpha_X$ is additive. We have already shown that this function is bounded and normalised. ■

Note that the above argument doesn't make use of Lemma 3.1.7 and hence it doesn't require Distribution Invariance.

3.1.4. Paris' Dutch book argument

As I said, the above proof is non-standard. It is quite different from most other treatments of the DBA, so I shall spend some time comparing it to another proof. I shall discuss J.B. Paris' approach to the Dutch book theorem. This part is based on Paris (1994, 2005 [2001]).

Paris considers the following scenario.

> [S]uppose that $0 \leqslant p \leqslant 1$ and that the expert is required to make a choice, for stake $S > 0$ between:
>
> - gaining $S(1-p)$ if $\mathbf{v}(X) = 1$ whilst losing Sp if $\mathbf{v}(X) = 0$
> - losing $S(1-p)$ if $\mathbf{v}(X) = 1$ whilst gaining Sp if $\mathbf{v}(X) = 0$
>
> Paris (1994, p. 20)

This is clearly a choice between a bet on X and a bet against X. In my above discussion of Halpern I just set the stake to be $S = 1$. If we do the same here, then we can see that Paris' framework relies on choosing between complementary bets. Paris defines a number which is supposed to be the maximum value at which you prefer the first of these to the second. This is exactly the definition of α_X we had above. The conclusion is, of course, that if the belief function defined by these maximum values is not a probability function, then you are open to a Dutch book: a set of bets that guarantees that you lose money.

It should be obvious that the overall value of a collection of bets of the above form is:

$$\sum_i S_i(\mathbf{v}(X_i) - p_i) \tag{3.2}$$

where taking the first kind of bet for X_i means setting S_i positive, and taking the second means setting S_i negative. Paris' definition of a Dutch book is a set of S_i and X_i such that the value of (3.2) is negative. He proves that there exists no Dutch book if and only if the p_is are convex combinations of $\mathbf{v} \in \mathbf{V}$.

It is implicit in his argument that one should always have a preference between complementary bets, which is exactly the content of (Complementarity). Paris simply does not worry about the possibility that the expert has no preference between the two bets. Paris makes this commitment explicit in his later paper.[15] Paris says that throughout the paper he is making the assumption that "for any $\theta \in SL$, stake $S > 0$ and $\eta \in [0,1]$ *either* one is willing to accept [the first bet] or one is willing to reverse roles and [accept the second bet]."

However, Paris is interested in a different kind of modification to the scenario. He is interested in valuation functions $\mathbf{v}$ that don't obey the classical rules.[16] And one of the situations he explains in his 2005 [2001] leads to the conclusion that belief functions should be Dempster-Shafer belief functions.

To reiterate: the valuation function $\mathbf{v}$ serves to decide which bets win, and which lose. So if you bet on an event X and it turns out that $\mathbf{v}(X) = 1$, you win. In the original Dutch book theorem, the $\mathbf{v}$ functions of interest are those whose outputs are classical truth values, ones that obey the classical truth conditions which we saw earlier in Definition 2.2.1.

Paris extends the Dutch book argument to cases where the $\mathbf{v}$ function need not obey the classical definition. Instead of learning the truth value of all atoms, you learn some (possibly non-atomic) proposition X. The $\mathbf{v}$ rule of interest now is $\mathbf{v}_X$ which operates like this: $\mathbf{v}_X(Y) = 1$ iff $X \models Y$. In the context of these alternative $\mathbf{v}$ functions, the "Dutch book method" then produces the result that to avoid Dutch books, your belief function should be a Dempster-Shafer belief function. In fact, Paris' result is quite general. For any set of valuation functions $\mathbf{W}$, Paris' theorem shows that to avoid a Dutch book you have to have a belief in the convex hull of $\mathbf{W}$. Paris shows that for a particular choice of $\mathbf{W}$, then using a result from Jaffray (1989), $\mathbf{W}^+$ is exactly the DS belief functions. So it is important how the "truth" of events is determined.

This is an interesting convergence of results. Above I showed that by modifying the restrictions on the agent's betting preference, you can derive a "Dutch book-like" result for DS theory. Paris shows that a similar result obtains if one instead

[15] Paris (2005 [2001], fn 2, p. 3)
[16] Classical truth valuations were defined in Definition 2.2.1 on p. 19.

keeps the betting preference restrictions in place, but modifies the conditions under which bets win. I think these two results are pulling in the same direction and it is the aim of section 3.1.6 to draw out these conclusions. This result is worth mentioning both because it gives an alternative route to nonprobabilism, but also because Robbie Williams has recently used a similar technique to subvert a different argument for probabilism (Williams 2012a,b).

3.1.5. Loose ends

There are some minor issues I would like to discuss in this short section. First, I want to address the difference between a fair bet and a favourable bet. And second, I wish to point out the difference between preference among bets and acceptance of a bet. Finally, I discuss the "converse theorem challenge".

Fair or favourable

Hájek (2008) rightly argues that the Dutch book argument is often set out in terms of "fair" bets: bets at odds that are fair according to your beliefs. This is to be contrasted with "fair-or-favourable" odds, which are bets at fair odds or better, by the lights of your beliefs. For example, say you were betting on a fair coin's landing heads. Your belief in this event is $\mathbf{b}(H) = \frac{1}{2}$. The odds of 2-for-1 on this event are fair by the lights of your belief. Longer odds are not fair, but favourable. If you're willing to accept 2-for-1 on heads, then of course it makes sense to take 3-for-1 or 10-for-1 on the same event (although you might get suspicious of someone offering 10-for-1 on a fair coin's landing heads). These bets are favourable. It is worth noting that Halpern's framework always talks in terms of fair-or-favourable bets and the norms govern both. α_X represents your fair betting quotient, anything less than that counts as favourable. Lemma 3.1.4 guarantees that favourable bets are preferred to their complements; Lemma 3.1.2 and Lemma 3.1.3 guarantee that more favourable bets are more preferred.

Preference or acceptance

Note that Paris' argument is in terms of bets you would *accept*, given your beliefs. Halpern's argument is only in terms of your *preference* among bets. To illustrate the difference between preference and acceptance, consider the fact that you can have a preference among bets without being willing to accept either. It seems rational to have the following preference:[17] $(\bot, 0.1) \succeq (\bot, 0.3)$, although you would certainly not want to buy either bet. That is, you can a preference for paying less for a bet, even if you don't want to buy any bet on an impossible event.

So how do we bridge the divide between preference among bets and acceptance of bets? We could characterise acceptance as follows: you accept a bet (X, α) just in case $(X, \alpha) \succeq (\neg X, 1 - \alpha)$. Or perhaps instead we could say that you prefer a bet just in case you prefer it to the neutral bet $(\top, 1)$. Or we could do this by appeal to $\mathbf{b}$. You would accept the bet (X, α) when $\mathbf{b}(X) \geqslant \alpha$. It should be obvious that $\mathbf{b}(X)$ will coincide with α_X by definition. These characterisations amount to the same thing.

Converse theorems

What is still missing from the above argument is the oft-neglected "converse theorem" (Hájek 2008). That is, we have not proved that if your betting quotients *are* probabilistic, then there does not exist a Dutch book against you. The above argument (in section 3.1.3) shows only that it is irrational to be nonprobabilistic. It does not show that being probabilistic protects you from this irrationality. In the present set up, proving the converse theorem would mean showing that if $\mathbf{b} \in \mathbf{V}^+$ then if you set your betting quotients in accordance with you $\mathbf{b}$, you satisfy Halpern's conditions. In fact, it would require you to have your preferences among bets determined *only* through the expectation associated with $\mathbf{b}$, in order to have you satisfy Halpern's conditions. This is possible, but I won't go through it here. Paris (2005 [2001]) is also immune to Hájek's "converse theorem" challenge. Paris explicitly makes his Dutch book argument an "if and only if" claim: so he effectively shows the converse theorem to be true as well.

3.1.6. What's wrong with the DBA?

This section mainly deals with why one might want to "get off the bus" before we add (Complementarity). That is, we are concerned with whether it might be reasonable to finish with just Dempster-Shafer betting quotients:[18] under what circumstances it might be reasonable to deny that one's betting preference should be constrained by (Complementarity). Reasons for this idea have to do

[17] Indeed, it follows from (Bet Dominance).

[18] I don't think DS theory is all that much better as a model of belief, but I do think it is a more reasonable restriction on rational betting quotients.

with concerns about vagueness or ambiguity undermining the rationale for the symmetry condition.

Consider an analogy to discussions of vagueness.[19] Imagine we are trying to determine the boundary between two regions. The boundary is vague. On one side, we are unequivocally in one region, call it X. On the other side we are certainly in $\neg X$. The boundary between them is indistinct. We can say that "up to at least this point we are definitely in X" without committing ourselves to saying "after this point we are certainly in $\neg X$." Now consider throwing a dart at this region (imagine that the dart has a uniform probability over the whole region). You are offered a bet on whether the dart will land in the X region. The nature of the game forces us to assign some numerical betting quotient to X, call it $\mathbf{b}(X)$. But because of the uncertainty, the vagueness, we are not thereby committing ourselves to saying that all the remaining probability belongs to $\neg X$, that is we are not saying that $1 - \mathbf{b}(X) = \mathbf{b}(\neg X)$. So you can recognise some bets on whether X will occur as being good, and some as being bad. But there is a third category: bets you aren't sure about.

Given that (Complementarity) seems to be relating belief in an event to belief in its complement, it seems like this would be the constraint to relax if one were worried about vagueness. I think the analogous move makes sense for cases of severe uncertainty or ambiguity. Paris' discussion of "the Dutch book method" for non-classical valuation functions (Paris 2005 [2001]) is also revealing. Here we seem to be saying that if the criteria for a successful bet are somehow "non-classical" then we need not restrict ourselves to probability measures. Vagueness might induce this kind of non-classical criteria for success.

The Paris approach seems to invite questions about the interpretation of the valuation function. Is this supposed to have some metaphysical import? What does it mean to say that the valuation function is non-classical? Does this mean that the world is somehow non-classical? It may be possible to give this valuation function an epistemic reading. However, the "deny (Complementarity)" approach I took in section 3.1.3 is, I feel, safer, since the rationality conditions on betting preference clearly have no metaphysical import. They are more clearly about the agent's attitude to vagueness. So my approach works even when we know that bet winnings are determined classically. Since my interest is in how to formally deal

[19] Fuzzy logic, one approach to dealing with vagueness, can also be modified to offer a theory of upper and lower probabilities: see Halpern (2003, pp. 40–3). So this analogy is not so far-fetched

with uncertainty and ambiguity, situating the problem in the agent's epistemic state is preferable to situating it in the logic of which bets win. Nevertheless, that Paris' approach leads to similar conclusions only strengthens the argument.

So nonprobabilistic betting behaviour can demonstrate, not your irrationality, but the weakness of your epistemic state. For example, Ellsberg style choices[20] do not show that you are behaving irrationally, but that you are in the possession of an incomplete, vague, ambiguous picture of the gambling scenario.

3.2. Representation theorems

The Dutch book argument is one big strand of argument for probabilism. The other such strand is the *representation theorem* argument. Representation theorems in slogan form say "If you conform to some reasonable restrictions on preference, and assumptions about the structure of the world; then there is a probability function that agrees with your choices." This is a very rough characterisation, there are myriad different representation theorems that are importantly and interestingly different. I look at some of these in turn, I will not discuss proofs in detail.

3.2.1. Representation in general

Before going on to talk about the representation theorems of specific interest in this context, I want to take some time to discuss the concept of representation in general. There is a mathematical field known as "measurement theory" which deals with these sorts of questions. We are ultimately interested in when one particular formal structure can be represented by another. In particular, we are interested in when certain kinds of order structures can be represented by a function.

I will briefly outline the simplest such case: length. Certain kinds of rigid bodies – sticks, pencils, book spines, arm spans – can be ordered by how long they are. For my purposes I take those things that can be so ordered to be called "sticks". Sticks needn't be stick-like, one can compare the diagonal size of a computer screen to the long edge of a piece of paper. The important thing is that there is an obvious way to compare these *lengths*. I briefly set out the basic idea of a representation theorem here.

[20] I discuss these games in the next chapter.

1. Length is a quantity – a property that admits of degrees – that attaches itself to some kinds of things.

2. Call "sticks" things that have a length. For my purposes, pencils, arms, book spines, the imaginary line between your outstretched fingers: these are all sticks.

3. Some sticks are longer than others. Say "$X \succeq_{\mathbf{len}} Y$" means "$X$ is at least as long as Y".

4. There is an operation you can perform on sticks: composition. You can lay sticks end to end and parallel. Call the compound of X and Y, $X \oplus Y$. It is also a stick.

5. The set of sticks ($\mathbb{S}$) has some structure. For all X, Z we have $X \oplus Z \succeq_{\mathbf{len}} X$. If $X \succeq_{\mathbf{len}} Y$ then $X \oplus Z \succeq_{\mathbf{len}} Y \oplus Z$.

6. There is a privileged stick: the null stick. No stick is shorter than the null stick.

7. Given some technical conditions, there is an additive function $\mathbf{len} \colon \mathbb{S} \to \mathbb{R}$ that assigns to each stick, a real numbered value: its length. **len** represents $\succeq_{\mathbf{len}}$ and is unique up to affine transformation.

8. By "**len** *represents* $\succeq_{\mathbf{len}}$" we mean that $X \succeq_{\mathbf{len}} Y$ if and only if $\mathbf{len}(X) \geqslant \mathbf{len}(Y)$.

9. By "**len** is additive" we mean $\mathbf{len}(X) + \mathbf{len}(Y) = \mathbf{len}(X \oplus Y)$.

Measurement theory studies this idea of representing a quantity. You get theorems that look like this: "If $\mathbb{S}$, $\oplus$ and $\succeq_{\mathbf{len}}$ have the right sort of properties, then the function **len** will have certain other properties." The properties of the function that represents the quantity tell us things about that quantity itself. For example, contrast length and temperature. Length has an additive representation. However, not so for temperature. There's no interesting physical composition procedure such that temperature is additive with respect to that procedure.[21] Put two thermal bodies in contact and the temperature of the composite body will be some sort of average of the two temperatures of the composite bodies before composition, not their sum. So this tells us that length and temperature are interestingly different *as quantities*, not just as regards their representations.

[21] Ellis calls length an *extensive* quantity, and temperature *intensive*.

There is a mathematical theorem that backs up point 7 on this list.[22] For the technical details, the classic work is Krantz et al. (1971). See also Ellis (1966) and Kyburg (1984) for more philosophical treatment.

I want to take a moment here to emphasise that only some aspects of the representation function can be understood to be telling us things about the world. For example, there exists a particular representation function that measures the particular stick X as having a length $\mathbf{len}(X) = 42$. This is a fact about the function that we don't take seriously as a fact about the world. The X stick has no intrinsic property of "forty-two-ness". The point is that there will be other functions that don't give X that value of length that will represent the quantity just as well. So a stick that is, say 42 centimetres will be some other number (16.5) of inches long. But it is the same length. So it is only properties invariant under an appropriate class of transformations that we take to tell us something about the world (Stevens 1946).

That, in brief, is what a representation theorem looks like. We will be interested in representation theorems that say something, not about length or temperature, but about your beliefs and values. The general case is taken up in Krantz et al. (1971, Chapter 5).

3.2.2. Eliciting beliefs

In the case of representations of subjective probability and utility, it is trickier than for length, because we are trying to represent two things – your beliefs and your values – but we only have the one yardstick: your choice behaviour. We take choice behaviour to reflect preference.

To get at the intuition behind representation theorems, I consider first ways you might try to elicit someone's degrees of belief in various events. Ramsey, De Finetti and Savage all thought about how one might go about finding out what an agent's actual degrees of belief might be through finding out what preferences they held. These theorems, being behaviourist in character, don't want to assume anything about your mental life. They look only at the overt, observable consequences of whatever goes on in your head. That is, they look only at your observed choices, which reveal your preferences among the options.[23]

What we effectively do in the Dutch book argument is use a given linear scale

[22] See Krantz et al. (1971, Chapter 3).

[23] I think this behaviourist character of the theorems has misled some people to think that only what can be elicited in this way can be real.

(money) to calibrate beliefs by preferences on bets. That is, we calibrate "strength of preference" on a previously given linear scale. This is like using a ruler to determine what our **len** function should return for a given stick. Or imagine learning that a column of mercury in a tube of constant diameter has a height proportional to the air temperature. We can then use our understanding of length (or height) to calibrate the temperature scale. Calibrating belief scales works in the same way. We can imagine offering you a series of bets where the event is fixed but the stake changes. There will be some "tipping point" value at which you no longer prefer the bet to its complement. We write this value down and repeat, thereby building up a record of your betting quotients (and therefore, it is assumed, your beliefs). We need a monetary scale – a ruler, if you like – to get this strategy off the ground. We are calibrating our belief scale (temperature) to the previously given linear scale of money (length).

The von Neumann–Morgenstern approach uses objective chances to calibrate utility (von Neumann and Morgenstern [1944] 2004). Instead of keeping the event fixed and shifting around the stake to elicit your beliefs, these authors use a different trick. They shift around the objective chances with which various consequences arise and calibrate their value function that way. Anscombe and Aumann (1963) use a similar trick to von Neumann and Morgenstern, but they use some objective chances to elicit subjective probabilities for other events as well.

Ramsey uses a neat trick to elicit strength of belief without recourse to such an explicit yardstick (Ramsey 1926). He assumes that there exist some propositions about whose truth value you have no preference either way. He calls these *ethically neutral* propositions. His idea is to pinpoint the "probability one half" point by finding an ethically neutral proposition $X_{0.5}$ with the following property: for outcomes $o_1 \succ o_0$, you are indifferent between gaining o_1 if $X_{0.5}$, o_0 if $\neg X_{0.5}$ and gaining o_0 if $X_{0.5}$, o_1 if $\neg X_{0.5}$. That is, despite preferring to get o_1, you have no preference about which of $X_{0.5}$ or $\neg X_{0.5}$ would be the better way to get it. Ramsey can then use these ethically neutral probability one half propositions to calibrate a utility scale. Choose o_0, o_1 such that $o_1 \succ o_0$. Now fix the utility of these outcomes as follows: $u(o_0) = 0, u(o_1) = 1$. There is an outcome, $o_{0.5}$ such that you are indifferent between it and o_1 if $X_{0.5}$, o_0 if $\neg X_{0.5}$. Fix $u(o_{0.5}) = 0.5$. Using a similar procedure you can find further outcomes that have utility 0.25 and 0.75, and so on.[24] To extend it to things preferred to o_1, we look for that

[24]We obviously need quite strong richness assumptions to guarantee that there are outcomes with

o_2 such that you are indifferent between o_2 if X, o_0 if $\neg X$ and o_1. Likewise for things dispreferred to o_0. Once Ramsey has built up this utility scale, it can be used to calibrate degrees of belief in a similar way to how we used bets above to calibrate belief. See Bradley (2004) for more details on Ramsey's method. What's impressive about Ramsey's approach is that no previously existing linear scale is needed to get the whole thing to work. As Bradley points out, however, there are some strong assumptions in the background. Ramsey's assumptions make his method a good way of *eliciting* beliefs, but the axioms don't make so much sense as constraints on rational preference.

Savage's method, like Ramsey's, doesn't need a previously existing scale to calibrate measures of belief against. Unlike Ramsey, Savage tries to make his axioms, as much as possible, reasonable constraints on rational preference. The order is also reversed: Savage derives a measure of belief first and uses that to fix the utilities.

3.2.3. Savage's Foundations of Statistics

Savage's representation theorem for personal probability is spread through chapters 2,3 and 4 of his classic *Foundations of Statistics* (Savage 1972 [1954]). Much like the Dutch book theorem as I presented it, the basic idea is to look at your choices or preferences under uncertainty and then to impute beliefs to you. Savage's approach is interesting, because he starts without making any assumptions about your utility scale. All he needs to assume is that you have preferences over a remarkably rich collection of acts. The trick is to help himself to particular kinds of acts, that can be used to elicit your beliefs about events. He builds a "qualitative probability" ordering out of the preferences over these sorts of acts, and then uses essentially the same proof as Krantz et al. (1971) to arrive at a probabilistic representation.[25] Once he has defined probability, he then goes on to show that the preferences determine a utility scale up to linear transformation.

Savage's world

Savage considers a set of possible worlds – the states – and defines *events* as sets of such worlds much as Halpern (2003) does. We can go the other way and define the events as the sentences of a language SL, and define the states as the atoms of this language. Alternatively, since there is a bijection between the atoms and

the right properties for all the values we want to specify.

[25]This is anachronistic. Krantz et al. acknowledge that their proof owes much to Savage's work.

members of **V**, we could have the states be described by the valuation functions. Some of Savage's conditions will require that there be uncountably many states.

Savage then defines consequences as those things that might matter to a person and that might be brought about by some act. The set of consequences **O** – what we earlier called outcomes – is effectively an unstructured set of things you might care about. It derives its structure from other features of the system. An act is then a function from atoms in SL to consequences **O**. I will denote the set of acts **A** as before.

An act determines, for each way the world could be, the consequences that would befall you were you to perform that action when the world were in that state. Finally, it is shown that given certain assumptions about the structure of preference among acts, there is one and only one probability measure that accords with preference in a certain way.

The restrictions on preference

In this section, I list the assumptions Savage makes about preference among acts. Reading "$a_1 \succeq a_2$" as "a_1 is at least as good as a_2", the first condition says that the relation is a total order.

P1 *For all* $a_1, a_2, a_3 \in \mathbf{A}$*:*

- $a_1 \succeq a_2$ *or* $a_2 \succeq a_1$
- *If* $a_1 \succeq a_2$ *and* $a_2 \succeq a_3$ *then* $a_1 \succeq a_3$

In other words, $\succeq$ satisfies transitivity and completeness. Note this is stronger than (Complementarity) from section 3.1.2, and similar worries apply here. Savage contemplates the possibility of allowing "$\succeq$" to be a partial order, but dismisses the project since doing so would be "losing much in power and advancing little, if at all, in realism" (Savage 1972 [1954], p. 21). In his presentation, Joyce makes the two parts of this definition distinct axioms. This is because the second part is arguably a principle of rationality while the first part is not (Joyce 1999, p. 84).

Savage defines various relations out of this preference relation. He needs the notion of "preference among acts, *given a certain condition is known to obtain*". That is, he needs preference conditioned on an event: "$a_1 \succeq a_2$ given X". For example, you may on balance decide you prefer to not carry an umbrella. However, your preference *given that it is raining* is to carry the umbrella. This is done by saying that if a_1 and a_2 lead to the same consequences outside of X, then $a_1 \succeq a_2$ given X

if and only if $a_1 \succeq a_2$. The second condition – known in the literature as the "Sure Thing Principle" – is then:

P2 *For every event X, the relation $\succeq$ given X is total.*

Calling this the STP is perhaps a little misleading: a lot of the work is being done by the definition of the relation "$a \succeq b$ given X". We will meet the STP again a number of times, and when we do meet it again, it will be more obvious what the content of the principle is. For now, I pass over it without further comment.

A constant act is one that gives the same consequence for all states. For all practical purposes, a constant act can be identified with its constant consequence. Savage relies on these constant acts to impose a preference ordering on the consequences through the ordering among acts. P3 says that for acts that are constant on X, whichever constant consequence is preferred determines which act is preferred given X.

P3 *If $a_1(x) = o_1$ and $a_2(x) = o_2$ for every $x \in [X]$, then $a_1 \succeq a_2$ given X, if and only if $o_1 \succeq o_2$*

Note that "$o_1 \succeq o_2$" is a preference between outcomes rather than acts. Savage determines preference between outcomes by the proxy of constant acts. Imagine there's an act that always gives outcome o_1. And another that always gives o_2. It's reasonable to understand preference between the outcomes as meaning a preference between the constant acts that give those outcomes.

An important assumption is being introduced here that is worth making more apparent. What this requires is that how much o_i is valued does not depend on which event obtains. We mentioned this kind of assumption earlier: how much you value the lottery that wins an umbrella depends on whether or not it rains. P3 rules out these sorts of possibilities. In many cases, the outcomes are not *state independent* in this way. For example imagine taking bets on future catastrophic climate change. The money you stand to gain if the climate system does go catastrophically wrong will likely be worth a lot less than the money is worth if things carry on as before. In this case, I'd recommend taking gambles that involve payouts in canned food, drinking water and solar panels. Or imagine making a big bet on the proposition "there will be hyperinflation in the UK by 2050". Of course, if you win, there will have been hyperinflation, and thus the money has become near-worthless. These examples illustrate state-dependent utilities. Having your values depend on the states like this is ruled out by P3.

The Savage-style response to these problems is to refine the outcomes. He suggests that – to return to the simple umbrella case – the outcomes have been mis-specified. The outcome is not in fact the same whether it's raining or not: the outcomes are "umbrella when raining" and "umbrella when sunny". With outcomes specified at that level of detail, the *Rectangular Field Assumption* (RFA) looks less convincing.[26] The RFA says that all possible functions from states to outcomes define acts. The more fine-grained you make these sorts of outcomes, the less reasonable it seems that you can find constant acts. For instance, what sort of act could possibly win you "umbrella when raining" when the event "it is sunny" obtains?

Savage then moves on to constructing a relation among events. Consider the choice between two acts: a_X, a_Y such that a_X wins you a prize when X obtains, and a_Y wins you the same prize when Y obtains. That is, $a_X(x) = o_1$ for all $x \in [X]$ and $a_X(x) = o_2$ for all $x \notin [X]$, and $o_1 \succeq o_2$ and similarly for a_Y.

	$\mathbf{v}(X) = 1$	$\mathbf{v}(X) = 0$
a_X	o_1	o_2
b_X	o'_1	o'_2

	$\mathbf{v}(Y) = 1$	$\mathbf{v}(Y) = 0$
a_Y	o_1	o_2
b_Y	o'_1	o'_2

Table 3.2.: Table of consequences for acts relating to P4

Whatever the prize is (assuming it is actually preferable to not receiving any prize) it seems reasonable that given the choice, you will pick the act that corresponds to the event you consider more likely (compare Lemma 3.1.3). In this way I could elicit your "qualitative probability" by asking you to pick between acts. Savage's P4 effectively says that for any two pairs of "prize functions" like this, you always pick the same way. Basically, this guarantees that the actual prize offered isn't relevant to which of the acts you pick.

P4 *For $o_1, o_2, o'_1, o'_2 \in \mathbf{O}$ with $o_1 \succeq o_2$ and $o'_1 \succeq o'_2$, and $a_X, a_Y, b_X, b_Y \in \mathbf{A}$ with $X, Y \in SL$: If the consequences for the acts are as in Table 3.2, then: if $a_X \succeq a_Y$ then $b_X \succeq b_Y$*

To rule out a particularly uninteresting pathological case, we assume that there are two consequences, such that one is strictly preferred to the other. This guarantees that there will be some nontrivial prize function.

[26]The term is due to Broome (1991, p. 80–1,115–7).

P5 *There are at least two constant acts,* o_1, o_2 *such that* $o_2 \not\succeq o_1$

Next, P6 claims that you can cut up the states of the world small enough such that changing the action of two acts on some small part of the space makes no difference to the preference order among those acts. This is basically a demand that the set of states be rich enough to be infinitely subdivided.

P6 *If* $a_1 \succ a_2$ *then, for every* $o \in \mathbf{O}$ *there is a partition of* SL *such that* $a'_1 \succ a'_2$ *and* a'_i *is identical to* a_i *on all but one partition element, and* a'_i *takes the value* o *on that one element.*

What P6 demands is that no matter how fantastic some consequence o is, you can always cut the event space up small enough so that an act modified to give you o on some tiny part of the space doesn't affect you preference among the acts. No matter how brilliant a prize, one can contemplate a situation so implausible that getting that prize in that implausible situation makes no difference to your preferences.

Finally, Savage's last constraint demands that if a is preferred to every consequence that b can produce on X, then $a \succeq b$ given X. It is, in short, a dominance condition.

P7 *If* $a \succeq b(x)$ *given* X *for every* $x \in [X]$*, then* $a \succeq b$ *given* X*.*

Note that "$a \succeq b(x)$" is actually a preference for an act over an outcome. As above, we understand this as shorthand for a preference for a over the constant act that always returns $b(x)$ for a fixed x. P7 has the flavour of what Bradley (2007) calls "The Averaging Slogan":

> No prospect is better (or worse) than its best (worse) realisation in a set of mutually exclusive and exhaustive prospects. (p. 241)

Savage proves that you can extract a "qualitative probability" from preference among acts. Savage shows that, given a rich enough space of states, acts and consequences, only one probability measure "agrees with" the qualitative probability thus determined. In the next section I will discuss qualitative probability, but first I shall mention some problems with Savage's approach.

Problems with Savage's approach

Here I outline a number of problems with Savage's approach.

Preference among *acts* is taken as the basic preference relation, and preference among consequences is derived from this. Psychologically, this seems backwards: it is because of a particular preference among the possible consequences — along with my judgements of the likelihoods of various possible states of the world — that I prefer some act to another. That is, it feels to me as if it is my beliefs and values that determine my choices, not the other way around. One might respond that yes, there is something psychologically backwards about this approach, but that's not necessarily a point against in as regards its adequacy as a model of belief. In the same way we can measure temperature by its effect on the height of a column of mercury, even though it is the temperature that determines the height, not vice versa. That is, the representation, the elicitation is focusing on the effect rather than the cause of the behaviour, but it is a meaningful representation nevertheless.

State independence – the outcomes' being independent of the states – seems to be a rather strong assumption. In some circumstances it seems reasonable: betting under normal circumstances where the economic conditions are assumed to be stable, for example. But in many real-world decision examples, the state of the world does impact the value of many of the outcomes we are trying to bring about. This is at least an assumption that can be dropped. There is a lot of work on "state-dependent utilities" and representation theorems for them (Schervish, Seidenfeld, and Kadane 1990). Given that many decisions based on scientific evidence, especially climate decisions, seem to have state-dependent utilities, it seems like this should be an assumption I should avoid. However, I think the question of state independence is orthogonal to my interest, which is in *severe uncertainty*. Given that this issue is difficult enough, I feel I am warranted in making the state independence assumption when I discuss decision making.

There is another kind of independence built into Savage's framework. This one is harder to get rid of, since it is not just an assumption that can be dropped. This is *act independence*. Act independence means that which act is performed does not affect the chances of the various states obtaining. For example, choosing to bet on *Categorical Imperative* does not make it more or less likely that that horse should win. This is built into the setup of Savage's system by having acts be functions from states to outcomes. Jeffrey's system, by doing away with this way of framing acts, doesn't build in act independence.

Act independence is interestingly and radically violated in some decision problems we face in a climate change decision context. Indeed, many of the options on the table are *specifically designed to lower the chance of catastrophic warming*. Any kind of mitigation action is a deliberate attempt to change the chances with which various climate states will obtain.

Act independence is a much more problematic assumption, I feel, than is state independence, in this context. However, for the most part I will be making the assumption. Again, this is not because I think it is a warranted assumption, but because I feel that the problems it raises are somewhat orthogonal to the problems of severe uncertainty. Even if the set of decision problems to which my conclusions are relevant is restricted to adaptation decisions (as opposed to mitigation decisions), I think a thorough understanding of severe uncertainty will be useful. It would certainly be interesting to drop this assumption, but I will have to leave this for future work. One approach would be to understand the states as being conjunctions of dependency hypotheses that tell you what would happen were a certain act to be performed. This seems to go against the colloquial understanding of what it is we have beliefs about: I have a belief about whether it will rain, not whether a certain bundle of dependency hypotheses ("If I were not take an umbrella I would get wet"...) holds. Even given this, it seems that such bundles of dependency hypotheses make the RFA less plausible in much the same way the "reconceptualise the outcomes" response to state independence does.

The richness of Savage's world – with its infinitely divisible state space – may seem troubling. For all practical purposes, only a finite number of elicitations can be done. So one can never in practice narrow down the representation of a real subject's mental state to unique probability and utility functions. But the more preferences are elicited, the more I can know about your beliefs and utilities. The richness is only really required for the uniqueness of the result, and can be seen as a kind of "limiting ideal" rather than as any requirement of rationality. I return to this point later.

Meacham and Weisberg (2011) object to the use of representation theorems as a basis for decision theory. Their argument is that representation theorems can't accurately characterise real agents' beliefs, nor can they provide a normatively compelling grounding. The failure of what they call *Characterisational Representationalism* relates to things I have already discussed: failures of real agents to satisfy the axioms (STP or completeness for example). The failure of *Normative Representationalism* is connected to an argument due to Zynda (2000) that any agent whose belief can be represented by a probability function can also be represented by a nonprobabilistic function, as long as a commensurate modification to the choice rule is also made. The upshot, for both Meacham and Weisberg and for Zynda is that we can't use representation theorems to get beyond qualitative belief orderings in terms of normative assessment of belief.

A stronger problem, perhaps, and one that I think Zynda's argument suggests is that "being *representable* some way or other is cheap. . . ; it's more demanding to actually *be* that way." (Hájek 2008, p. 803). Hájek goes on to give the example of *Voodooism* as a rival to probabilism. You prefer A to B just in case there are more A-favouring spirits than there are B-favouring spirits. The claim is that agents can be *represented* as having such warring voodoo spirits, but that does not make it the case that there are such spirits.

3.2.4. Representing qualitative belief

Whether you get your qualitative belief ordering from Savage's elicitation procedure, or just try to justify the axioms on $\succeq_{\mathbf{b}}$ directly, you end up at the same point.

Savage's theorem works by using your preferences to determine a *qualitative belief ranking*. Fine (1973) just starts with the qualitative belief order and shows it is representable by a probability function if it satisfies certain axioms. These axioms are similar to the ones that Savage imputes to you on the basis of your preferences over acts. Krantz et al. (1971) give a similar result. There is much work in this area on minimal necessary and sufficient conditions for a probabilistic representation.

Ultimately, all the arguments require that the ordering be complete. They also all require that the domain of objects of belief be uncountably rich. These are necessary conditions on the probabilistic representation's being unique. If the ordering is only partial, or if the space of beliefs is only finite, then there will be many probability functions that represent the ordering. The idea is that either of these weakenings introduces "wiggle room" that allows one to shift the probability mass around between propositions and maintain the same ordering among the elements. When the ordering is dense, this wiggle room is severely constrained.

Joyce (1999, pp. 91–92) sets out the axioms more or less as follows:[27]

- Normalisation: $\top \succ_{\mathbf{b}} \bot$

[27] I have suppressed those details that only matter for the case of countable additivity. Happily this also means that the alternative axiomatisations of Krantz et al. (1971) or Fine (1973) coincide perfectly for this part.

- Boundedness: $\top \succeq_{\mathbf{b}} X \succeq_{\mathbf{b}} \bot$ for all X
- Ranking: $\succeq_{\mathbf{b}}$ is a partial order
- Completeness: $X \succ_{\mathbf{b}} Y$ or $Y \succeq_{\mathbf{b}} X$ for all X, Y
- Quasi-additivity: If $X \wedge Z \equiv \bot \equiv Y \wedge Z$ then
  - $X \succ_{\mathbf{b}} Y$ iff $X \vee Z \succ_{\mathbf{b}} Y \vee Z$
  - $X \succeq_{\mathbf{b}} Y$ iff $X \vee Z \succeq_{\mathbf{b}} Y \vee Z$
- Richness: If $X \succ_{\mathbf{b}} Y$ then there exists a partition of SL: $\{Z_i\}$ such that $X \succ_{\mathbf{b}} Y \vee Z_i$ for all i

The first couple of these look very similar to conditions we saw imposed on belief functions in section 2.3.2. Some of the conditions should also remind us of conditions from Savage's axioms for preference. Qualitative belief that satisfies the above axioms can be represented by a unique probability function. Proofs of this result can be found in Fine (1973); Krantz et al. (1971); Villegas (1964).

Note agents with DS beliefs can violate Quasi-additivity. Let $\mathbf{bel}(Y) = 0 = \mathbf{bel}(Z)$ and let $\mathbf{bel}(X) > 0$. Let $Z = \neg Y$. Note that $\mathbf{bel}(X \vee Z) \geqslant \mathbf{bel}(X)$ but otherwise it is unconstrained; so set $\mathbf{bel}(X \vee Z) < 1$. However, $\mathbf{bel}(Y \vee Z) = 1$ because DS belief is normalised. This violates quasi-additivity. The trick is that a lot of "belief mass" can accrue to disjunctions, while leaving very little mass on the individual disjuncts. So if one thought that Quasi-Additivity captured something distinctive about the structure of qualtiative belief, and thus of degree of belief more generally, then DS belief would not be a good representation of it.

3.2.5. Other representation theorems

Savage's wasn't the first representation theorem, nor was it the last. I don't aim to be exhaustive in my survey of representation theorems, but I do want to mention some other important theorems.

The first is Richard Jeffrey's representation theorem, building on mathematical work by Ethan Bolker. This is an interestingly different theorem, because it does away with some of the constraints imposed by Savage's framework. This comes at the cost of the representation not being unique.

The second representation theorem I discuss is due to Luce and Raiffa. This is important because they use *choice sets* rather than preference relations as the basic entity. I will make use of this way of speaking when I discuss imprecise decision.

Jeffrey's system

I won't go into much detail about Jeffrey's system.[28] As I mentioned in chapter 2, Jeffrey makes states, outcomes and acts all into propositions. So he doesn't build in the act-independence of the states as Savage does. Nor does he make the same strong state-independence of the outcomes assumption that Savage does.

These differences make his representation theorem much harder. In fact, the representation he achieves is not unique in the way Savage's is. One can get a unique representation only by extending Jeffrey's framework. There are two ways to do this. One way, due to Joyce (1999), is to augment the set up with a second relation: a relation of qualitative belief. That is, you have a preference relation and a qualitative belief relation, and these – under certain assumptions – give you a unique representation. The other way of achieving uniqueness, due to Bradley (1998), is to turn the algebra into a *conditional algebra* by adding a particular kind of (non-truth-functional) conditional binary operator to the language.

Despite being Savage-like in taking acts to be functions, my approach is influenced by Jeffrey and by Joyce's discussion of Jeffrey because I am aware of the importance of the assumptions I am making about what acts are. Indeed, one of the important lessons learned from studying imprecise decision is that the details of how acts bring about outcomes is important. This idea is even more important in *causal decision theory*, where you explicitly model ideas about how your actions influence the outcomes.

Luce and Raiffa's choice set approach

There is an interestingly different approach to representation of belief in Luce and Raiffa (1989). Instead of representing a preference relation, Luce and Raiffa model an agent as having a *choice set*. That is, they model an agent as choosing some subset of the options available as "the good ones". The conclusion that is drawn from their discussion is that if you satisfy the conditions on your choice set, then there is a probability function such that your choice maximises expectation with respect to it.

I mention this because later (in section 5.1) I will use the related idea of a "choice function" as my primary representational tool when discussing imprecise decision. A choice function, $\mathcal{C}$, is a function from a set of acts into a subset of that set: $\mathcal{C}(A)$ picks out the best options from A.

[28]But see Jeffrey (1983) and chapter 4 of Joyce (1999).

Luce and Raiffa impose rationality constraints on the choice set such that there is a preference relation that can be closely associated with it. By this I mean that the conditions on the choice set are such that something is in the choice set if and only if it is maximal with respect to the associated ordering.[29]

Informally, their important axioms are as follows (see Luce and Raiffa 1989, pp. 286–93):

- $\mathcal{C}(A)$ is never empty
- If $f \in \mathcal{C}(A)$ and g dominates f (in the sense that $g(X) \geq f(X)$ for all X) then g is in $\mathcal{C}(A)$
- If some g in A dominates f, then f is not in $\mathcal{C}(A)$
- Adding a new act to the set of options can't make an old non-choiceworthy act choiceworthy, and can only make a previously choiceworthy act non-choiceworthy if the new act is itself choiceworthy
- The Sure Thing Principle
- Probability mixtures of choiceworthy acts are choiceworthy (i.e. the choice set is convex)

They state a theorem (p. 293) to the effect that if your choice rule satisfies these properties, then there is some probability function such that you can be represented as maximising expected utility with respect to it. Luce and Raiffa note that it is not the case that every act that maximises with respect to that probability need be such that your choice rule deems it choiceworthy. Satisfying Luce and Raiffa's conditions makes maximising expectation a necessary condition on choiceworthiness, but not necessarily a sufficient condition.

3.2.6. Richness, Completeness and Extendability

Common to all the above arguments is an assumption that whatever relation you are dealing with – a preference relation or a qualitative belief relation – it should be complete, and it should be defined on a suitably rich algebra of events.

The richness and completeness assumptions are not required to demonstrate the *existence* of a probability representation, but only to prove its uniqueness. In the next chapter I will argue that *sets* of probabilities are a better representation

[29]The results are due to Amartya Sen (Sen 1970, 1977).

of uncertainty, so I am happy with non-unique representation. Indeed, I will later argue that non-unique probabilistic representation is one way to justify imprecise probabilism (section 4.4.1).

Let's think how we might defend these assumptions. Let's take the richness first. One might argue that instead of being a requirement on actual beliefs, the richness assumption could be an assumption about *extending* your beliefs. The idea being that it should be possible for you to continue adding more propositions to your event algebra indefinitely. Arguably, this understanding of the richness assumption as one of extendability is the understanding Savage himself had of it. He says:

> Suppose, for example, that you yourself consider $B \prec_{\mathbf{b}} C$, that is, you would definitely rather stake a gain in your fortune on C than on B. Consider the partition of your own world into 2^n events each of which corresponds to a particular sequence of n heads and tails, thrown by yourself, with a coin of your own choosing. It seems to me that you could easily choose such a coin and choose n sufficiently large so that you would continue to prefer to stake your gain on C, rather than on the union of B and any particular sequence of n heads and tails.
>
> Savage (1972 [1954], pp. 38–9, with some notational change)

So Savage thought that arbitrarily fine distinctions of probability judgement could be made by extending the algebra to include a suitably rich collection of otherwise uninteresting and independent events.

Some have suggested that completeness is also like this (Hawthorne 2009; Jeffrey 1983; Joyce 1999). Hawthorne for example says that incomplete belief orderings are perfectly reasonable. However, it would be unreasonable for an agent to have a preference that could not be extended into a complete one. He says that there being no complete extension of your belief ordering points to an implicit incoherence. That is, if X and Y are incomparable, and there is no complete extension of the relation, then neither of $X \succeq Y, Y \succeq X$ is even possible for you. This seems to be incoherent. So Hawthorne requires only that the relation be completeable. He does the same sort of thing for richness.

You might fail to have a preference for several different reasons. For example, you might fail to have a preference simply by never having considered the possibility of having to make the choice between the options.[30] This is incompleteness

[30] I will talk about preference, but I feel that much the same argument can be made for the qualitative belief ordering.

through failure of introspection. Or perhaps you fail to have a preference simply because the evidence you have about the choices is so weak that no preference is sanctioned. This is a stronger kind of failure of completeness. Some might want to argue that only the first kind of incompleteness is permissible: only through failure of introspection can your preferences be incomplete. I would argue that the second kind of incompleteness is also permissible. There can be cases where even when you turn your attention to the choices, you remain unable to determine a preference either way.

3.2.7. Against behaviourism

In both the Dutch book argument and the representation theorem sections, I have suggested that I find completeness of preference to be an unreasonable axiom. There is, however, one school of thought that would take issue with this claim: behaviourism.

The behaviourist might argue against incomplete preferences as follows: having no preference should be modelled as being indifferent. There's no way to elicit the difference between the two states, so there is no real difference between them. The claim is that there is no behavioural difference between the two attitudes. That is to say that no (reasonable, robust) method of preference elicitation can distinguish these two kinds of attitude.[31] So, the argument goes, we should not represent differences we cannot elicit. My response to this is just to say that there is more to belief than dispositions to choose. So possibility of elicitation isn't an appropriate necessary criterion for the reality of some aspect of belief. Mark Kaplan argues as follows:

> Both when you are indifferent between *A* and *B* and when you are undecided between *A* and *B* you can be said not to prefer either state of affairs to the other. Nonetheless, indifference and indecision are distinct. When you are indifferent between *A* and *B*, your failure to prefer one to the other is born of a determination that they are equally preferable. When you are undecided, your failure to prefer one to the other is born of no such determination. Kaplan (1996, p. 5)

If pressed to choose, you will plump for one or other of the incommensurable options. A behaviourist could take this fact to point to some sort of disposition to act. This disposition to act in this way when faced with that choice is exactly

[31] Ittay Nissan has helpfully pressed me on this point.

what the preference relation is supposed to be capturing. The behaviourist would argue that this suggests there really is no difference. I would respond to this by arguing that there is an important distinction in the character of the dispositions in the incommensurability case as opposed to the preference case: a distinction worth representing.

On a purely instrumentalist level, we can still show that there is some difference in character between indifference and incommensurability. If you are indifferent between a and b, then you will prefer $a + \varepsilon$ to b for any small positive gain ε. You cannot induce a preference between incommensurable goods by this procedure of "sweetening" one of the options. So the behaviourist is wrong to suggest that there is no difference to elicit between indifference and incommensurability.

The more general point to make as regards my disagreement with the behaviourist position is the following. Behaviourists take probability and utility to be merely a summary of a consistent agent's preferences. The summary can then form a useful tool to predicting further action: one could always work out what your preferences are given the preferences you have already demonstrated, but using the summary functions makes things easier. I don't find behaviourism at all compelling as an explication of belief. Or rather, I take your elicited behaviour to be only part of what characterises your belief state. I however take probability and utility to be independently meaningful psychological quantities of some form or another. Instead of summarising preference, I take probability and utility (or rather, belief and value) to be the *inputs* to a decision problem. It is the job of decision theory to help you manipulate those inputs in such a way as to arrive at the best outcome. If we took the behaviourist's view on probability and utility, the action guiding aspect wouldn't involve decision theory at all: the action guiding advice from a behaviourist amounts to "introspect on your preferences and choose the most preferred option". To me this is backwards: we are using what we know about how you value things and how likely you think things are to *determine* what the preferences should be. The preference determining procedure should determine preferences that are consistent, that is, that are in line with the rationality axioms of decision theory. So there is some feedback between preference determining decision procedures and preference constraining axioms of rationality. And the representation theorem would tell us that any decision procedure that determines preferences that satisfy these axioms is effectively a maximising expected utility procedure. *Would*, that is, if all the axioms of the representation theorem had independent intuitive force as constraints on preference. I have argued that they don't and that the behaviourists' response to my criticism misses the point, since I am coming at decision theory from a different angle.

Bermúdez (2009) distinguishes several different projects that use decision theory as a tool. These are decision theory used to guide action; to normatively assess action; and to explain and predict action. He also outlines two understandings of utility: the substantive and the operational. The substantive understanding of utility takes utility to be a measure of some sort of "goodness": some aspect of the outcomes that feeds into the motivations for choice. The operational understanding of utility takes utility to be part of the behaviourist preference-summarising machinery I discussed above. I situate myself as being primarily interested in the action-guiding and normative dimensions of decision theory, and endorsing a substantive understanding of utility. The behaviourist takes the operational understanding of utility and is more interested in the explanation or prediction dimension of decision theory. Given these differences of outlook, it is not surprising that we should disagree.

Joyce goes further and argues that not only are the two outlooks distinct, but also that the behaviourist outlook is deficient.

> There are just too many things worth saying that cannot be said within the confines of strict behaviourism... The basic difficulty here is that it is impossible to distinguish contexts in which an agent's behaviour really does reveal what she wants from contexts in which it does not without appealing to additional facts about her mental state... An even more serious shortcoming is behaviourism's inability to make any sense of *rationalizing explanations* of choice behaviour. Joyce (1999, p. 21)

Ultimately, the behaviourist's "explanation" of your choice is that you were satisfying your preferences. This is unhelpfully circular.

3.3. Epistemic utility

Dutch book arguments and representation theorems have been criticised for relying on agents' actual or hypothetical choices. The pragmatic nature of these arguments has been questioned (Christensen 2001, 2004). It has been suggested that preference among acts is not the sort of thing that should constrain epistemic rationality. These pragmatic considerations, it is claimed, should not influence the structure of your beliefs. To counter this criticism, some authors have offered arguments for probabilism that rely on purely epistemic premises. Pettigrew (2011) outlines three types of epistemic utility argument. I will outline the general approach and introduce the several types of epistemic utility argument. I will then outline the premises of one specific argument, that of Joyce (1998). The proof will be relegated to Appendix A.

3.3.1. Scoring rules in general

The idea common to all these types of argument is to score a belief function, and the aim is to have beliefs that score well. For example, one sort of score you might use is the "distance from the truth". If you had some appropriate sort of measure of distance from truth, then you could use this to score various kinds of beliefs. This is the kind of approach that Joyce takes. He gives a characterisation of what properties an *inaccuracy measure* should have and then proves a theorem: "If your credences are incoherent (nonprobabilistic), then there is a coherent (probabilistic) belief function that is more accurate in every possible world". This is a *domination* argument.[32] Another argument that builds on inaccuracy measures is that of, for example Leitgeb and Pettigrew (2010a,b). Instead of being accuracy dominated, in this argument it is shown that nonprobabilistic beliefs have higher *expected inaccuracy*. A slightly different argument is motivated by the idea that it is not closeness to the truth but calibration that is the important thing about beliefs. The proofs behind these types of arguments then show that probabilistic beliefs are better calibrated. Finally, there are arguments that start from the idea that whatever else is true of a good belief function, it should expect itself to be the best. That is, it is a flaw in a belief function if by its own lights, some other function does better. Scoring rules that reward this property of belief functions are known as *proper scoring rules.* These sorts of arguments (like Predd et al. (2009)) are called *propriety* arguments.

All the arguments have a similar structure. They all purport to show that probabilistic credences are better because they do better at some putative epistemic good-making feature.

[32] Domination arguments are not one of Pettigrew's categories. The remaining three types – from Pettigrew (2011) – are all kinds of *expected utility* arguments.

3.3.2. Admissible scoring rules

For now, I focus my attention on Joyce (1998). What is important about your beliefs is that they are accurate. If it were the case that your degrees of belief could be more accurate than they actually are, I would consider that a failure of rationality. Scoring rules arguments start from the axiom that it is desirable to minimise the inaccuracy of your credences. More specifically, they argue that it is a defect to be *accuracy dominated*. Call your current credences $\mathbf{b}$. If there is some other belief function, $\mathbf{b}'$ that is more accurate in every possible world, then your rationality is impugned. That is, if your beliefs are such that there is a way you could *always*[33] have more accurate beliefs, it's irrational to stick with the defective $\mathbf{b}$. The conclusion of scoring rules arguments is typically that you should be probabilistically coherent. That is, you should have degrees of belief that satisfy the axioms of probability.

Much hangs on exactly how one formalises the idea of inaccuracy, and Joyce criticises several earlier attempts at doing this for begging the question: by having the idea of probability somehow built in to what inaccuracy measure to use. Joyce's own approach is to impose some conditions on what counts as a legitimate scoring rule. The argument then goes "Whatever scoring rule is picked from this class, your beliefs are accuracy dominated unless you are probabilistically coherent". This result is rather surprising. It says that if your degrees of belief are not probabilistic, then there's a probability measure which does better *however the world turns out*.

This section explores Joyce's discussion of accuracy measures. This mainly draws on Joyce (1998). I will also mention criticism of Joyce's paper in for example Hájek (2008) and Maher (2002) and Joyce's response to these (Joyce 2009).

In his 1998 paper, Joyce seems to allow the event space SL to be finite or countable, though Maher (2002) and Joyce (2009) restrict themselves to the finite case. I shall stick to the finite case throughout.

Now we are ready to start characterising inaccuracy measures.

Definition 3.3.1 *An inaccuracy measure is a function* $\mathbf{I} : \mathbf{B} \times \mathbf{V} \to \mathbb{R}$.

$\mathbf{I}(\mathbf{b}, \mathbf{v})$ is a measure of how far from the "truth" (designated by $\mathbf{v}$) the belief function $\mathbf{b}$ is. So, Joyce's claim becomes the following: whatever $\mathbf{I}$ you are using, if your

[33] Actually, it says you always do at least as well and sometimes strictly better. In the interest of rhetorical force, I've swept aside some of the subtlety.

$\mathbf{b}$ is probabilistically incoherent (i.e. if $\mathbf{b} \in \mathbf{B} \backslash \mathbf{V}^+$) then there is $\mathbf{c} \in \mathbf{V}^+$ such that $\mathbf{I}(\mathbf{c}, \mathbf{v}) \leqslant \mathbf{I}(\mathbf{b}, \mathbf{v})$ for all $\mathbf{v}$.

A final point before we come to the conditions themselves is that Joyce makes extensive use of the idea of "mixtures" of belief functions. If $\mathbf{b}, \mathbf{b}' \in \mathbf{B}$, then there is a "line" between them. This geometrical structure is what Joyce exploits in the proof.

Definition 3.3.2 $\mathbf{b}\,\mathbf{b}'$ *is the "line segment" between* $\mathbf{b}$ *and* $\mathbf{b}'$. $\mathbf{c} \in \mathbf{b}\,\mathbf{b}'$ *if and only if* $\mathbf{c} = \lambda\,\mathbf{b} + (1-\lambda)\,\mathbf{b}'$, *for some* $\lambda \in [0,1]$. *That is:* $\mathbf{c}(X) = \lambda\,\mathbf{b}(X) + (1-\lambda)\,\mathbf{b}'(X)$ *for all* X.

With these preliminaries in place it is time to look at the conditions that Joyce imposes on inaccuracy measures.

> Structure *For each* $\mathbf{v} \in \mathbf{V}$, $\mathbf{I}(\mathbf{b}, \mathbf{v})$ *is a non-negative, continuous function of* $\mathbf{b}$ *that goes to infinity in the limit as* $\mathbf{b}(X)$ *goes to infinity for any* $X \in SL$.

Inaccuracy $\mathbf{I}(\mathbf{b}, \mathbf{v})$ clearly attains a minimum of 0 when $\mathbf{b} = \mathbf{v}$. So $\mathbf{I}$ is non-negative. This minimum being zero is partly conventional. But it would seem unusual to say that a belief function has some non-zero inaccuracy with respect to itself... That is, if you fully believe all the truths and fully disbelieve all the falsehoods, it would be strange to insist that your beliefs had any inaccuracy other than zero. By tending to infinity in the limit, Joyce is merely trying to capture the idea that, since $\mathbf{v}$ is bounded, so should $\mathbf{b}$ be. Patrick Maher offers an alternative description of this last part of the Structure condition.

> Structure (Maher) *For all* $X \in SL$ *and* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that, for all* $\mathbf{b} \in \mathbf{B}$, *if* $|\mathbf{b}(X)| > \delta$ *then* $\mathbf{I}(\mathbf{b}, \mathbf{v}) > \varepsilon$ *(Maher 2002, p. 78)*

Since the biggest a truth value can get is 1, as your belief function assigns bigger and bigger numbers to X, the more inaccurate it must get. Note this is due to the convention that valuation functions assign the value 1 to the truth. But this isn't doing much work. We only need that truth and falsity are assigned bounded values. You could replace 0 and 1 throughout with $\mathbf{v}(\bot)$ and $\mathbf{v}(\top)$ and nothing drastic would happen to the argument.

> Extensionality *At each possible world* $\mathbf{v}$, $\mathbf{I}(\mathbf{b}, \mathbf{v})$ *is a function of nothing other than the truth values that* $\mathbf{v}$ *assigns to propositions in* SL *and the degrees of confidence that* $\mathbf{b}$ *assigns to these propositions.*

With this condition, Joyce simply wants to rule out anything other than the truth and the belief affecting inaccuracy. For example, simplicity, verisimilitude, pragmatic usefulness of beliefs are all barred from directly affecting the "goodness" of a belief. This differentiates his project from, for example that of Maher (1993), who is interested in a broader notion of epistemic utility.

Joyce argues that the norm of accuracy is by far the most important norm when one is concerned with epistemology. One could also argue that by having accurate beliefs, you also help yourself conform to some lesser norms. For example, having accurate beliefs will normally be pragmatically advantageous. Of course one can construct cases where believing the truth would be pragmatically bad, but in general, the two aims will line up.

Joyce also argues that having a false but "almost true" belief (a *verisimilar* belief) will lead one to have other true beliefs.[34] So while both Copernicus and Kepler had false beliefs about the orbit of the Earth, Kepler's beliefs would be overall better. Copernicus believed the orbit of the Earth around the Sun was circular. Kepler believed the orbit was elliptic. The actual orbit is almost elliptic, but perturbed by the influence of the Moon and other celestial bodies. So both Kepler and Copernicus held a false belief to a high degree, but one would like to be able to say that Kepler's beliefs were *somehow* better, in virtue of being closer to the truth. Joyce's response to this objection is to say that despite being false, Kepler's belief in an elliptic orbit around the Sun would lead him to have other, strictly true beliefs that Copernicus didn't have. Joyce gives the example of "that the average distance of the Earth to the Sun is different in different seasons". So "the overall effect of Kepler's inaccurate belief was to improve the *extensional* accuracy of his system of beliefs as a whole." (p. 592 Joyce 1998, Joyce's emphasis). Strictly speaking if one's beliefs differed *only* in what one believed about the shape of Earth's orbit, then they would be equally inaccurate. Joyce's point is that believing the more verisimilar proposition will cause changes in other beliefs which will be strictly true. These differences cannot be captured within the gradational accuracy framework. But then, these differences rely on the semantic content of the propositions about which we have beliefs, so the inability of the framework to fully capture these things is unsurprising and should not be seen as a defect. In short, the claim is that verisimilar beliefs have more true consequences than less verisimilar beliefs. A proper argument for this claim would involve developing a method for counting the consequences of sentences.[35] In any case, Joyce (2009)

[34]Discussions with Foad Dizadji-Bahmani helped clarify this point.

[35]This might be tricky. If V is the more verisimilar belief and V' the less verisimilar one, then

doesn't rely on extensionality.

And now, onto the next condition.

> Dominance *If* $\mathbf{b}(Y) = \mathbf{b}'(Y)$ *for every* $Y \in SL$ *other than* X *then* $\mathbf{I}(\mathbf{b},\mathbf{v}) > \mathbf{I}(\mathbf{b}',\mathbf{v})$ *if and only if* $|\mathbf{v}(X) - \mathbf{b}(X)| > |\mathbf{v}(X) - \mathbf{b}'(X)|$

What Joyce is getting at with this condition is that inaccuracy of a belief function somehow supervenes on the inaccuracies of the individual beliefs that make it up. Those individual inaccuracies are straightforwardly measured by the distance from the truth (designated by $\mathbf{v}(X)$).

> Normality *If* $|\mathbf{v}(X) - \mathbf{b}(X)| = |\mathbf{v}'(X) - \mathbf{b}'(X)|$ *for all* $X \in SL$*, then* $\mathbf{I}(\mathbf{b},\mathbf{v}) = \mathbf{I}(\mathbf{b}',\mathbf{v}')$

Again, this condition is saying that if some belief function is always exactly as far from the truth as another, then they are just as inaccurate as each other. It is quite a strong condition. It says not only that if $\mathbf{b}$ and $\mathbf{b}'$ are always equally far from $\mathbf{v}$, they are equally inaccurate, but further that even if they are the same distance from *different valuations* – $\mathbf{v}, \mathbf{v}'$ – they are equally inaccurate.

The final two conditions are the most controversial. They are the two that Maher (2002) criticises, and they are crucial to Joyce's result. They relate to the idea of mixtures of belief functions.

> Weak Convexity *Let* $\mathbf{m} = (1/2\,\mathbf{b} + 1/2\,\mathbf{b}')$ *be the midpoint of the line segment between* $\mathbf{b}$ *and* $\mathbf{b}'$*. If* $\mathbf{I}(\mathbf{b},\mathbf{v}) = \mathbf{I}(\mathbf{b}',\mathbf{v})$*, then it will always be the case that* $\mathbf{I}(\mathbf{b},\mathbf{v}) \geqslant \mathbf{I}(\mathbf{m},\mathbf{v})$ *with equality only if* $\mathbf{b} = \mathbf{b}'$.

Think of $\mathbf{b}$ and $\mathbf{b}'$ as two epistemic agents. What this condition says is that if the two of them are equally inaccurate, they can't do better than effecting some sort of epistemic compromise. This condition says roughly that if two epistemic peers disagree on something, they should "split the difference".[36] That is, if $\mathbf{b}$ is just as inaccurate as $\mathbf{b}'$ then an equal mixture of the two, $\mathbf{m}$ will be no more inaccurate.

The final condition says that equally biased mixtures will always be equally inaccurate:

> Symmetry *If* $\mathbf{I}(\mathbf{b},\mathbf{v}) = \mathbf{I}(\mathbf{b}',\mathbf{v})$*, then for any* $\lambda \in [0,1]$ *one has:*
>
> $$\mathbf{I}(\lambda\,\mathbf{b} + (1-\lambda)\,\mathbf{b}',\mathbf{v}) = \mathbf{I}((1-\lambda)\,\mathbf{b} + \lambda\,\mathbf{b}',\mathbf{v})$$

whenever X is a consequence of V, $V' \vee X$ is a consequence of V'.

[36] There has been much recent discussion on the correct response to disagreement among peers. Joyce's paper predates much of it. See Christensen (2009) for an overview.

Consider the case where λ is large. The left hand side of this equation is a mixture that is biased towards $\mathbf{b}$ and against $\mathbf{b}'$. The right hand side is biased towards $\mathbf{b}'$. What the condition says is that these two biased "compromises" between $\mathbf{b}$ and $\mathbf{b}'$ will be equally inaccurate. These two conditions impose constraints on how inaccuracy of mixtures is related to inaccuracy of the things they are mixtures of.

Joyce's main theorem, then, in its final form is as follows:

> (**Main Theorem**) If gradational inaccuracy is measured by a function $\mathbf{I}$ that satisfies Structure, Extensionality, Dominance, Normality, Weak Convexity and Symmetry, then for each $\mathbf{c} \in \mathbf{B} \backslash \mathbf{V}^+$ there is a $\mathbf{c}' \in \mathbf{V}^+$ such that $\mathbf{I}(\mathbf{c}, \mathbf{v}) > \mathbf{I}(\mathbf{c}', \mathbf{v})$ for *every* $\mathbf{v} \in \mathbf{V}$.
>
> Joyce (1998, pp. 597–8)

3.3.3. Comments on epistemic utility

There are a number of problems that have been raised with this argument. I think the same sort of problems crop up against the other types of epistemic utility argument, so while the following discussion will be largely in terms of Joyce's argument, things should carry over straightforwardly to the other arguments I mentioned earlier.

The underdetermination of the inaccuracy measure

The argument deliberately says as little as possible about *which* inaccuracy measure you should use: this makes the argument as general as possible. But it does lead to a problem: $\mathbf{I}$ is underdetermined. Which inaccuracy measure you use will affect *which* probability function accuracy dominates your nonprobabilistic belief. So without knowledge of which inaccuracy measure is "the right one" you don't know how to change your credences so as to avoid being accuracy dominated. In fact, there are two problems here. First, just knowing that your beliefs are accuracy dominated doesn't tell you how to change your beliefs. Knowing your beliefs are dominated tells you that they are bad, but it doesn't show that the dominating beliefs are good. Of course, the dominating beliefs are *better* in that they don't have the defect of being dominated, but there might be an even better belief to move to which doesn't dominate your current beliefs. So, even if you knew the right inaccuracy measure, the advice you would get from this domination argument would be fairly weak. And if we don't even know which inaccuracy measure we should use, we can't even make use of that weak advice.

Does it even make sense to talk about there being some correct measure of inaccuracy? And if not, what exactly is the normative upshot of Joyce's argument? Hájek (2008) worries about this. In short, his worry is that the theorem does not guarantee that moving to any probability function will make you less inaccurate.

> Here's an analogy (adapted from Aaron Bronfman and Jim Joyce, personal communication). Suppose that you want to live in the best city that you can, and you currently live in an American city. I tell you that for each American city, there is a better Australian city. (I happen to believe this.) It does not follow that you should move to Australia. If you do not know which Australian city or cities are better than yours, moving to Australia might be a backward step. You might choose Coober Pedy. Hájek (2008, p. 808, fn. 18)

Hájek then mitigates the worry as follows:

> That said, the calibration argument may still be probative, still diagnostic of a defect in an incoherent agent's credences. To be sure, she is left only with the general admonition to become coherent, without any advice on how specifically to do so. Nevertheless, the admonition is non-trivial. Compare: when an agent has inconsistent beliefs, logic may still be probative, still diagnostic of a defect in them. To be sure, she is left only with the general admonition to become [consistent], without any advice on how specifically to do so. Nevertheless, the admonition is non-trivial. (ibid)

Perhaps you could take the set of all probability functions that accuracy dominate your nonprobabilistic belief relative to some set of reasonable inaccuracy measures. This set of probabilities would then be the right belief to have. I will return to this in the next chapter.

Nonclassical logic and accuracy domination

Williams (2012a,b) points out that Joyce's argument is sensitive to the underlying logic in exactly the same way that Paris (2005 [2001]) showed Dutch book theorems are. That is, Joyce's result rests on proving something about the convex hull of the truth valuations. So in the same way the Dutch book theorem ends up sanctioning Dempster-Shafer betting quotients if the underlying logic is a kind of non-classical one, Joyce's argument leads to the same conclusions.

Not all epistemic utility arguments have this feature, however. For example, the theorem of Joyce (2009) is sufficiently different that the same trick doesn't obviously work. The assumption of classical logic comes in, in that paper when the set of "world vectors" is defined as the set of vectors with a 1 in one position and 0 everywhere else: this framework is built up from a partition of the event space. A nonclassical world vector might allow more than one atom to get a nonzero value. In any case, it isn't clear what the upshot of Williams' argument is when applied to these newer gradational accuracy arguments.

The size of belief space

It is worth noting that in setting up his framework, Joyce is thinking of **B** as a vector space. Maher picks up on this point and adds that Joyce is considering **B** to be a subset of $\mathbb{R}^n$ where n is the number of atoms. So vector $\overrightarrow{b}$ represents a belief function **b**, for some ordering of the atoms in the following way: the length of the projection of $\overrightarrow{b}$ onto the i^{th} coordinate gives the value of **b** for the i^{th} atom. What value does the vector representation give to compound sentences, for example $t_i \vee t_j$? There is a whole class of belief functions that agree on values for the atoms but disagree on compound sentences.[37] Leitgeb and Pettigrew (2010a,b) explicitly restrict themselves to this way of thinking. This focus on the atoms is a subtle way to beg the question. It shifts the goalposts in favour of those functions that are determined by their actions on the atoms only, that is, in favour of $\mathbf{V}^+$. Note that functions that are superadditive but not additive aren't determined by the action on the atoms, since superadditivity involves an inequality, rather than an equality.

For this reason one might prefer to think of the vector space in question being of dimension of the cardinality of SL. That is, there is an axis for *every* proposition, not just the atoms. There is then some non-trivial structure on the "world vectors". For example, if a particular "world vector" has 0 in the i^{th} proposition, and j^{th} proposition, then, if k is the proposition corresponding to the disjunction of the i^{th} and j^{th}, the k^{th} coordinate must be 0 too. But what determines this structure? Perhaps the structure of the **v** functions determines the structure of the world vectors. Now there should be some connection between a coordinate representing a disjunction and coordinates representing its disjuncts, but what should that relationship be? I don't claim to have an answer to this question, but I do want to stress that assuming that a belief function is determined by its action on the atoms is a substantive assumption. The same argument works for those arguments that

[37] Take for example, any Dempster-Shafer belief function that assigns 0 to all the atoms.

take the belief function to be defined over some partition of the space (Joyce 2009; Predd et al. 2009).

Real valued belief, real valued inaccuracy

There is a further point: Joyce takes his belief functions to be real valued functions. No other types of functions are even considered to be candidates for representing belief. As I will argue in the next chapter, there is a better representation in terms of *sets* of probability functions. Joyce also takes his inaccuracy measures to be real valued functions.

I think uncertainty is not a single-dimensional thing. There are importantly different ways to be uncertain about something. These differences cannot be represented by mapping propositions onto a linear, one-dimensional scale like the real line. What gradational accuracy arguments secure is the conditional claim that *if* belief and inaccuracy are to be modelled by real valued functions, *then* probability theory is privileged. I deny the antecedent of this claim. All the other gradational accuracy arguments I am aware of make the same assumptions and are thus open to the same criticism.

In chapter 6, I will discuss a case study of scientific uncertainty. The uncertainty we encounter there is certainly not single-dimensional.

The structure of belief space

As well as the size of belief space, there is another issue to do with its structure. Joyce's Structure condition included the claim that $\mathbf{I}(\mathbf{b}, \mathbf{v})$ is a continuous function for each $\mathbf{v}$. Continuity of a function is defined relative to a topology on its domain. The topology doesn't necessarily have to be specified directly: sometimes there will be some obvious metric on the space, and the metric induces a topology (Sutherland 1975). Joyce can't really be appealing to some implicit metric on the space, since his whole proof relies on constructing a measure of distance on the space of belief functions. Joyce is perhaps appealing to the standard topology on the vector space representation of the belief function. But the standard topology on a vector space is one induced by a standard metric.

The intuition behind demanding that $\mathbf{I}$ be continuous is this: "a tiny change in one of my beliefs, $\mathbf{b}(X)$ say, shouldn't cause a huge difference in the inaccuracy of my beliefs". The topology (or the metric) is what gives content to the idea of a "tiny change" in belief. So maybe what Joyce needs is some kind of proposition-wise continuity of inaccuracy relying just on differences between $|\mathbf{b}(X) - \mathbf{v}(X)|$.

This is in fact what Pettigrew (2012) has, instead of continuity.[38] But, even this use of absolute difference could be criticised for implicitly relying on the notion of Euclidean distance. Why should the proposition-wise distance of my belief from the truth be measured by the Euclidean distance between them? Why not $|\mathbf{b}(X) - \mathbf{v}(X)|^2$ or $\log|\mathbf{b}(X) - \mathbf{v}(X)|$?

It seems that I might be labouring this point, but the history of analysis is rife with examples of how the intuitive concept of "continuity" is a tricky beast to pin down formally. Appendix 1 of Lakatos (1976) gives a lively account of one such episode. Given that Joyce's "Lemma-1" (Theorem A.2.1 in the current setting) relies on a result from topology, it is important to understand what this continuity assumption means. It seems the implicit topological/metrical notions are doing some work for Joyce, so I think it is worthwhile voicing these worries about that framework. Leitgeb and Pettigrew (2010a) and its followup are much more explicit in the presuppositions they rely on, but the privileged role given to Euclidean notions of distance could still be questioned in their proof.[39]

Quite a bit of what Joyce wants to do with this idea of continuity can be achieved just by thinking about "continuity along mixtures". For example, in the proof of Theorem A.2.2, we only need continuity along the lines of mixtures. But Theorem A.2.1 relies on $\mathbf{V}^+$ being a closed, bounded region. These topological properties can't be captured as easily by just focusing our attention to lines. The theorem of Joyce (2009) also relies on the inaccuracy functions being continuous: the theorem involves the use of Brouwer's fixed point theorem. It would be interesting to see exactly what conditions need to be imposed on the space of belief functions. Assuming the standard topology on the reals and assuming Euclidean distance on the axes (on the propositions) is certainly sufficient. But are there weaker conditions for which the theorem holds? Or can an elementary justification for why the standard topology and Euclidean distance gives the right notion of "a tiny change in belief"?

There isn't time to discuss in detail other importantly different and interesting theorems in this area. There are a few niggling technical worries with the original theorem, only some of which are addressed in the literature since Joyce (1998). I think ultimately justifying probabilism through gradational accuracy is hamstrung by the insistence on real-valued belief and real-valued inaccuracy, and by

[38] Pettigrew's other conditions are slightly different from Joyce's, so it's simplifying things somewhat to say that his "Proposition-wise continuity" does all the work of Joyce's continuity assumption.

[39] Leitgeb and Pettigrew make this very point themselves

underdetermination of the inaccuracy measure.

3.4. Other arguments

3.4.1. Cox's theorem

A strangely neglected argument for probabilism can be found in a paper by physicist R.T. Cox from 1946. The main thrust of the argument is that we want our degrees of belief to match up with classical logic in the limit, and codifying this desideratum in a particular way forces us to conclude that degrees of belief must be represented by something isomorphic to a probability measure. Cox takes his cue here from Keynes' "logical theory" of probability (Keynes 1921). Probability is here taken to be a "logic of plausible inference" or "logic of reasonable expectation".

The basic entity in Cox's proof is "$(A|X)$" which is the degree of plausibility of proposition A given state of information X. And $(A|B,X)$ is interpreted as the degree of plausibility of A given the state of information X *plus* the truth of B. This is not quite how Cox himself expressed it, this is the reformulation of Van Horn (2003). Cox had the "X" also being a true proposition and read $(A|B)$. as "the degree of plausibility of A given the truth of B". The difference is subtle and I won't dwell on it here. For now, it suffices to say that a state of information can include true propositions, but can also contain "soft information that says nothing with certainty but still affects one's assignments of plausibilities" (Van Horn 2003). One might imagine that X could contain statistical information: this says nothing with certainty, but can still give you strong reason to believe certain generalisations.

So these things, $(A|X)$ and so on, are by definition, real numbers. They measure the degree to which some proposition is plausible given certain evidence. These numbers aren't completely unconstrained. For example, $(A|A,X)$ seems like it should have some maximal value: the degree of A's plausibility given A's truth should be fixed. True things are maximally plausible, or rather, things known to be true are maximally plausible. The plausibility of $A \wedge B$ is somehow related to the plausibility of its components and so on.

These kind of concerns are what drives Cox's argument. Cox's proof proceeds by giving some plausible constraints on how $(\neg A|X)$ should relate to $(A|X)$, for example. And how $(A \wedge B|X)$ should relate to $(A|X), (B|X), (B|A,X)\dots$ and so on. Given the functional relationships between these numbers, Cox proves that the only functions that will do the job are isomorphic to probabilities.

I follow the presentation of Van Horn (2003) in this section. Other treatments of the theorem, outside of Cox's 1946 paper, can be found in: Halpern (1999a,b); Jaynes (2003); Paris (1994).

Van Horn calls the requirements "R1,R2…"; I will replace these with more semantically transparent labels. So let's get on to the first requirement.

Basic $(A|X)$, *the plausibility of proposition A given state of information X is a single real number.*

This condition just tells you what kind of thing we are taking plausibilities to be. It is a substantive claim to say that plausibilities can be represented by a single real number: it implies that all plausibilities can be compared. I have already discussed my scepticism of this sort of demand for complete comparability. It rules out it ever being the case that there are two propositions such that you can't tell which is more plausible. I have already discussed my views on the reasonableness of incommensurability.

Maximality *There exists a real number* $\mathbb{T}$, *such that for every X and A,* $(A|X) \leqslant \mathbb{T}$

This condition says that there is some maximal value that plausibility can take. Nothing can be strictly more plausible than something known to be true. Van Horn uses "$\top$" for this value, rather than "$\mathbb{T}$", but I want to keep "$\top$" reserved for the *proposition* that gets the plausibility value "$\mathbb{T}$". At this stage I'd be happy to just set $\mathbb{T} = 1$ by fiat, as this convention fits with what I've done in earlier sections. However I refrain from doing this, because Cox actually *proves* that this must be the case.

Both the above two requirements are contained in Van Horn's R1, but I have separated them, because I think each deserves comment on its own.

Next, a bit of terminology. A state of information, X, is *consistent* if there is no proposition A for which both $(A|X) = \mathbb{T}$ and $(\neg A|X) = \mathbb{T}$. This is the "plausibility analogue" of a set of sentence's being consistent if there is no sentence such that it and its negation can be derived from the set.

Our next condition is in the same vein: it says that these plausibility values behave in the obvious way as regards logic.

Compatibility *Plausibility assignments are compatible with the propositional calculus:*

- *If A and A′ are tautologically equivalent, then* $(A|X) = (A'|X)$

- *If A is a tautology, then* $(A|X) = \mathbb{T}$
- $(A|B, C, X) = (A|B \wedge C, X)$
- *If X is consistent and* $(\neg A|X) < \mathbb{T}$ *then A, X is also consistent.*

This just says that plausibility behaves as we would expect vis-à-vis logic. So tautologically equivalent propositions are equally plausible, tautologies are maximally plausible. This is motivated by the desire that our logic of plausible inference be continuous with classical logic in the limit.

These have been fairly straightforward conditions so far. Now the requirements get more interesting.

Negation *There exists a non-increasing function S_0 such that* $(\neg A|X) = S_0(A|X)$ *for all A and X*

This is effectively saying that the plausibility of the negation of a proposition stands in some functional relationship to the plausibility of the proposition itself. The function is non-increasing because as A gets more plausible (as $(A|X)$ gets bigger) $\neg A$ shouldn't get more plausible as well.[40] Some of what I said against Complementarity will also apply here. I don't think that, in general, this tight link between an event and its negation should hold.

Note that $\models A \leftrightarrow \neg\neg A$ so by Compatibility, $(A|X) = (\neg\neg A|X)$. From this it follows that $(A|X) = S_0(S_0(A|X))$. If we define $\mathbb{F} = S_0(\mathbb{T})$ then we know that $\mathbb{F} \leqslant (A|X) \leqslant \mathbb{T}$. This is so since $(A|X) = S_0(\neg A|X) \geqslant S_0(\mathbb{T})$ by Maximality and the fact that S_0 is non-increasing. Van Horn says that this holds for consistent X, although I'm not sure the qualification is necessary.

So far, so uncontroversial. The next requirement is not so obviously reasonable, and much of the debate around Cox's theorem has been on whether or not this requirement is justifiable.

Universality *There exists a nonempty set of real numbers P_0 with the following two properties:*

- *P_0 is a dense subset of $(\mathbb{F}, \mathbb{T})$. That is, for every pair of real numbers $a, b \in P_0$ such that $\mathbb{F} \leqslant a < b \leqslant \mathbb{T}$ there is some $c \in P_0$ such that $a < c < b$.*
- *For every $y_1, y_2, y_3 \in P_0$ there exists some consistent X with a basis of at least three elementary letters – call them A_1, A_2, A_3 – such that* $(A_1|X) = y_1, (A_2|A_1, X) = y_2$ *and* $(A_3|A_1, A_2, X) = y_3$.

[40] The domain of S_0 is the subset of the real numbers that are plausibility values.

A *basis* for X is the set of elementary letters that X tells us about. That is, X is a state of information that is informative about the set of elementary letters in its basis.

This requirement feels pretty strong: it requires that the space of plausibilities be quite rich. There are no finite dense sets, so this requirement forces there to be infinitely many values that plausibility can take, and thus infinitely many propositions. Van Horn's name for this condition "Universality" hints at the reason why he thinks it isn't as problematic as it seems. The idea is that if this is supposed to be a logic of plausible inference – something that extends classical logic – then it should be similarly context neutral. It should apply in all situations. The idea is that whatever else our logic should do, it should "be capable of handling a case where we have three completely unrelated atomic propositions with arbitrary plausibilities." (Van Horn 2003, p. 12) It is this assumption that Halpern (1999a) criticises. Similar worries arise with Savage's assumptions about the richness of the state space (P6).[41]

It is reasonable to expect that the plausibility of the conjunction of propositions is somehow related to the plausibility of the propositions individually. In fact it turns out that $(A \wedge B|X)$ is related to certain conditional plausibilities as well. Van Horn (2003) discusses why this should be. It should not be wholly unintuitive that the following functional dependence should hold.

> Conjunction $(A \wedge B|X) = F((A|B, X), (B|X))$ *where F is a strictly increasing function in both arguments.*

This should perhaps remind you of the discussion of conditional probability and in particular of the (Conditional) condition (p. 32).

Jaynes (2003) gives the following reasoning for the above requirement:

> In order for $A \wedge B$ to be a true proposition, it is necessary that B is true. Thus the plausibility $(B|X)$ should be involved. In addition, if B is true, it is further necessary that A should be true; so that plausibility of $(A|B, X)$ is also needed... [If a] robot has $(B|X)$ and $(A|B, X)$ it will not need $(A|X)$. That would tell it nothing about $A \wedge B$ that it did not have already.
>
> Jaynes (2003, Following Van Horn in adapting the notation)

[41]But see Arnborg and Sjödin (2000) for a possible way out. The trick is to replace "infinitely rich algebra" with "algebra that could be extended so as to be infinitely rich". We have already encountered the idea of "richness as extendability" so I won't say any more here.

The idea is that $(A \wedge B|X)$ can't be worked out from just $(A|X)$ and $(B|X)$, since we don't know the correlation between them. But with $(A|B,X)$ and $(B|X)$, we know everything there is to know.

These requirements allow us to prove a variety of facts about our plausibilities which will ultimately force them to be isomorphic to probabilities. For example, the function F in our final condition is such that, for all x, y, z:

$$F(x, F(y, z)) = F(F(x, y), z) \tag{3.3}$$

This follows from the associativity of logical conjunction. A result from functional analysis shows that any function with this form is such that there is a continuous, strictly increasing function g with $g(F(x, y)) = g(x) + g(y)$. Defining $w(x) = \exp(g(x))$ allows us to say that:

$$w(A \wedge B|X) = w(A|B,X)w(B|X) \tag{3.4}$$

These things are now starting to look a lot like probabilities. In fact, one can prove that $w(\mathbb{T}) = 1$ and $w(\mathbb{F}) = 0$, which is strange, since these values are typically taken to be conventional.

Van Horn then defines α as the fixed point of S_1, a function that extends S_0 to all of $[\mathbb{F}, \mathbb{T}]$. That is, $S_1(\alpha) = \alpha$. Where S_1 is a function on $[\mathbb{F}, \mathbb{T}]$ such that for all A and consistent X, $S_1(A|X) = S_0(A|X)$. This α is unique since S_1 is non-increasing. Then r is defined as follows: $r = -\log 2 / \log w(\alpha)$. Finally, define $\mathbf{pr}(x) = w(x)^r$. This $\mathbf{pr}$ is then shown to be a probability measure. $\mathbf{pr}^{-1}$ is shown to be well defined, so there is a tight relationship between x and $\mathbf{pr}(x)$, so much so that whether we talk about the plausibilities x or their associated transformation into probabilities $\mathbf{pr}(x)$ makes no difference.

All this manoeuvring allows us to show that $\mathbf{pr}(\neg A|X) = 1 - \mathbf{pr}(A|X)$, which, together with the above results, shows that our plausibilities $(A|X)$ are isomorphic to probabilities through the mapping $\mathbf{pr}$.

This has been a fairly cursory look at Cox's theorem, but I imagine I am already trying my reader's patience by belabouring these points. The argument is mathematically sophisticated but philosophically unilluminating.

I will not spend long commenting on Cox's theorem, since my criticisms of it have already come up elsewhere. Having plausibility measured on a real valued scale is too strong as it rules out incommensurability. I have discussed this when talking about preference relations, and again when talking about Joyce's assumption that belief is measured with real valued functions. Second, the negation assumption builds in a strong complementarity between an event and its negation; a complementarity that I feel is unwarranted. I discussed much the same thing in regard to Complementarity in section 3.1.6. Jon Williamson offers several criticisms of this argument for probabilism. Among them, he suggests that the continuity of F – the conjunction function – "seems to be less a rational constraint than a technical convenience" and that the basic assumptions make too strong an assumption about an agent's logical omniscience (Williamson 2010, p. 34).

3.4.2. Structure via calibration

One might think that reflection on the objects of belief might suggest something about the structure of belief.[42] We're happy with the idea of using the logical structure of the objects of belief to justify (weak) norms of belief. Consider the earlier discussion of logical omniscience: if two sentences stand in a particular logical relationship, then you are committed to treating them the same way. Perhaps other aspects of the structure of the objects of belief can ground norms of belief. Can probabilism be grounded by investigating this structure?

Think about flipping coins or drawing cards: it seems like beliefs about these things are constrained in various ways. Our beliefs about coin flips should match up with the chances, or with the observed frequencies. This offers another way of arguing for probabilism. Paris (1994, pp. 17–9) shows that observed relative frequencies are probabilities. This suggests that if you want to apportion your belief to the evidence, you should at least have your beliefs be probabilistic. All the ways the evidence might be share this structural feature, so your belief should also have this feature.

We can see the argument that frequencies must be probabilistic as an instance of the theorem that probabilities are convex combinations of truth values (Theorem 2.3.2). It is thus open to the same trick as Paris (2005 [2001]) applies to the Dutch book argument and Williams (2012a,b) does to gradational accuracy arguments. Namely, changing the underlying logic affects the outcome. One way we might rationalise this move in the current context is as follows: Perhaps the statistics you have gathered are deficient in some way. Perhaps when counting people by hair colour, the light wasn't always so good, so you weren't sure whether someone's hair was brown or black. What do you do? Perhaps you guess: you say

[42] This section draws on my paper "Nonprobabilistic chance?" Thanks to audiences at the LSE PhD. student seminar and at the British Society for the Philosophy of Science annual conference 5–6 July 2012, University of Stirling.

"it might have been brown or black, let's say black". Or perhaps you'd reason as follows: "I can't tell whether Bob's hair was brown or black, so let's put him in the disjunctive 'brown or black' category, without putting him in either of the disjuncts." The statistics you build up in this way will determine a non-classical valuation function, and your statistics will be superadditive, but not additive. This is what Walley and Fine (1982) do, and what Walley (1991) does in much more detail. Or you might just argue that there is genuine indeterminacy or vagueness in the world. So what to conclude about what norms of belief are sanctioned by the structure of events is at the very least dependent on some substantive assumptions about the nature of events.

This argument also has the flaw that it only works for cases where there is statistical evidence of the right sort. We might cast our net further and argue that beliefs should match the objective chances or rather, your beliefs should at least have the structure common to all possible chances. And if the objective chances must be probabilistic, so too should the beliefs be. More carefully, there are various principles – like Lewis' Principal Principle for example – that suggest that beliefs are somehow constrained by chances. According to the PP, credences should be expectation values of chance functions. If chance functions are probabilities, then so should the credences be.

I have yet to see convincing arguments that chances must have the structure of probabilities. One would need such an argument to have the above line of reasoning hold any water. In fact, there are some reasons to think that they needn't be. For example, if objective chances are some kind of quantity – a property that admits of degrees – then there is some sort of thing to which the quantity attaches to. When we were discussing length earlier I called the things that the quantity "length" attaches to "sticks". Let's call the things chance attaches to "events". For chances to be probabilistic, events must have a certain structure: they must form an algebra. One thing to note is that we need $X \wedge Y = Y \wedge X$, since we must have $\mathbf{ch}(X \wedge Y) = \mathbf{ch}(Y \wedge X)$. Typically in formal representations of quantum events, like in quantum logic, you don't have this commutativity. It is true that in all observable bases, you do have commutativity, and indeed the mod-squared amplitudes are additive, but this isn't generally true (Rédei 1998; Rédei and Summers 2007). Krantz et al. (1971) discuss "QM-qualitative probabilities", where these differ by not always having conjunctions. That is, it can be that X and Y are in your event structure, but $X \wedge Y$ isn't.[43] Note this is

[43]For instance, if X says something very specific about a particle's position and Y something

a different point from "Humphreys' paradox" (Humphreys 1985; Suárez 2007, forthcoming). The issue there is what the link is between conditional chances and conditional probabilities.

There is also a well-known tension between the demands of probability theory and infinite spaces. Consider throwing a dart at the real unit interval. What is the chance that the dart hits a rational number? The intuitive answer is that $\mathbb{Q} \succeq_{\mathbf{ch}} \bot$: that is, hitting a rational number is more likely than the impossible event. This is a long-winded way of saying that hitting a rational number is *possible*. But probability theory forces upon us the conclusion that $\mathbf{ch}(\mathbb{Q}) = 0$. So it seems that probabilistic chance isn't making some distinctions we might want to make between these sorts of events.[44] There has been some back-and-forth on hyperreal probability theory recently that I don't want to get into, but it is another thing to bear in mind (Hájek ms.).

In any case, it seems like intuitions or folk conceptions of the "event concept" aren't strong enough to support the claim that chances are probabilities.

3.5. The interaction between the arguments

I have outlined several arguments for probabilism. How are we to assess these arguments? What criteria are appropriate to an assessment of arguments for probabilism?

3.5.1. Criteria for assessing the arguments

A first possible way of distinguishing between the arguments is in terms of the size and structure of the domains the argument requires. For example, does it only work for finite spaces of events? Perhaps you are a finitist, and find arguments that require infinite spaces unappealing. Perhaps you think that given the richness of the structure of the actual world, arguments that only hold for finite domains are worthless. In either case, this is a relevant variable for assessment.

Another yardstick against which to measure different arguments for probabilism is how well they deal with conditional probabilities. We have independent reasons

specific about its momentum, then while $\mathbf{ch}(X)$ and $\mathbf{ch}(Y)$ might have values, $\mathbf{ch}(X \wedge Y)$ doesn't make sense, since it would violate the uncertainty principle. See p. 214 of Krantz et al. (1971).

[44] That is, there is a tension if you buy into the idea that $\mathbf{ch}(X) > 0$ is a necessary condition for X being possible. Hájek and Schaffer (2007) both endorse this idea.

for thinking that conditional probability should be taken as the basic notion and standard unconditional probabilities defined in terms of them (see, for example Hájek 2003). So it seems that it would be a point in its favour if one of these arguments could be straightforwardly applied to conditional probabilities as well as to unconditional ones. I am talking here about synchronic conditional probabilities as the basic notion: this should not be confused with discussions of updating one's probabilities on learning new evidence via conditionalisation. These two ideas are obviously related, but they are not the same thing.

On the topic of conditionalisation, we could ask about other norms that are closely associated with probabilism. For example, can there be an analogous argument for the norm that you should update your beliefs by conditionalising? Or can you justify a norm of calibration this way? It seems like a good epistemology should have its norms justified in harmonious ways.

You might also have a particular view on the nature of belief, and thus find certain arguments appealing or unappealing depending on whether they fit with your views. For example, many people think that belief is an epistemic thing, and thus a prudential argument like the Dutch book argument is getting at the wrong thing.

Finally, it seems like a flaw in an argument if it is easy to modify to give different conclusions. That is, if "nearby" arguments suggest different norms of belief, then this undermines the correctness of the original argument. For example, as mentioned earlier, J.B. Paris gives a Dutch book argument for DS belief functions by modifying the valuation functions. So a probabilist who wants to use Dutch book arguments must really be committed to the valuations' being classical to secure her conclusion. Then the argument is only as good as her argument for the claim that "ideal" bets are settled by classical logic.

3.5.2. Assessing the arguments

Now I wish to assess each of these on each of the criteria I mentioned above. The first of these was the nature of the objects of belief. More specifically, I am interested in the structure and the size of the space of propositions. Probability measures are defined over an algebra of events or propositions. What's at stake here is what sort of assumptions are made about this algebra in the above arguments. Gradational accuracy arguments typically assume that the algebra in question is finite. Joyce (1998) doesn't make clear whether he restricts himself to finite domains, but Maher (2002) takes him to be doing so, and Joyce (2009) is explicit about it. Dutch book arguments normally tend to assume finiteness as well. There are results for infinite cases though they are significantly harder. It also seems like any pragmatic argument that requires you to take infinitely many bets is significantly weaker. The jury is still out on whether countable additivity is reasonable. Howson (2009) argues that it is not, while Williamson (2010) takes it to be so.

Representation theorems, on the other hand, tend to assume that the space of acts over which one is choosing is really quite rich and this makes the space of beliefs very rich as well. That is, it is not only infinite, but structured in a particular way.[45] Fishburn (1973) says that typically one assumes that the set of states is non-denumerably infinite. Finiteness does not preclude there being a probabilistic representation, but it does preclude its uniqueness. Cox's theorem also requires that the algebra of propositions be structured in a way that can only be satisfied by infinite spaces. Halpern (1999a,b); Paris (1994) make these claims but see Snow (1998) for a dissenting opinion.

Which of these (finite or rich) you find more appealing depends on your taste. On the one hand it seems that your boundedness, your finitude means that it is reasonable to assume you have only a finite number of opinions. On the other hand, it certainly feels as if the world is rich enough that the things about which you could have opinions is unbounded, potentially infinite.

Now, which arguments work with conditional probabilities? Cox's theorem is already couched in terms of conditional probabilities, so that's a point in its favour. Fishburn (1973) offers a representation theorem that grounds conditional probabilities in terms of conditional acts.

There are arguments to the effect that Dutch book arguments cannot be used to justify conditional probability. Döring (2000) argues this point.[46] As for gradational accuracy, it is an open question, as far as I know. That is, there are no results I am familiar with in this vein.

Now, which arguments can be used to offer other important norms for belief like conditionalisation, calibration, equivocation? Dutch book theorems do well here: there are arguments of a similar character for all of the above norms. For example, Paris (1994, pp. 22–23) or Halpern (2003, pp. 77–78). For the latter two – calibration and equivocation – see, for example, Williamson (2010). Leitgeb and Pettigrew (2010a,b) offer a gradational accuracy argument for conditionalisation,

[45] For example, Fine's axioms C5 and C6 (Fine 1973)

[46] Döring seems to suggest that Jeffrey conditionalisation is immune to his criticism, but that would mean building a whole new foundation of Jeffrey conditional probabilities

but as Pettigrew (2011) says, there are many open questions in this area still. Representation theorems for updating look very much like representation theorems for conditional probabilities, but with the extra axioms being understood in a different way.

Which arguments fit with which conceptions of *what beliefs are*? If you are inclined to think that beliefs are the sort of things that are action guiding, then it seems that you will be convinced by Dutch book-type arguments. If, on the other hand, beliefs are something you ascribe to others in order to understand their actions, then representation theorems seem like the right way to go. If beliefs are somehow internal but not necessarily action guiding – if belief is an attitude we take towards propositions – then gradational accuracy arguments, or qualitative belief representation look like the most reasonable kinds of argument. Cox's theorem looks like it is motivated by a similar conception of beliefs to gradational accuracy arguments, but rather than looking at the relationship between propositions and the world, it looks at the relationships of partial entailment among the propositions themselves.

Related to this last criterion is the question of whether utility is important. Dutch books and representation theorems seem to be tied up in practical rationality and questions of action, reward, punishment and so on. Gradational accuracy or Cox's theorem, on the other hand, seem to be oblivious to questions of utility. If you think your belief shouldn't be tied to your conception of what is practically good, then the first two kinds of arguments are not going to be convincing. Having said that, one might take the Dutch book to be merely dramatising an inconsistency in a nonprobabilistic agent's credences.[47] This would suggest that the focus on utility is misleading and that Dutch book arguments aren't so grounded in pragmatic concerns. Gibbard (2007) goes further in arguing that this is something of a false dichotomy. He argues that a pure concern for truth can only be a concern for the guidance value of the beliefs. So even purely epistemic values are somewhat instrumental.

One might ask: are there arguments similar to the ones above but whose conclusion is something other than probability? That is, is there a Dutch book theorem that justifies, not probability, but Dempster-Shafer belief theory, for example? My discussion of the Dutch book theorem highlights exactly these sorts of results. The relevance of this is as follows: say you need four assumptions about betting behaviour in order to prove the Dutch book theorem. Now, imagine that in the

[47] The phrase is from Hájek (2008).

absence of one of those conditions, you can prove instead that degrees of belief ought to be Dempster-Shafer functions. This makes you wonder whether that fourth condition is reasonable, and whether it's better to stick with DS theory. In a way, these other theorems "in the neighbourhood" of the Dutch book theorem, weaken it, or at least make you scrutinise the assumptions. The theorem would come as a package, each of its presuppositions integral to the argument.

Williams (2012b) has an argument based on the method in Paris (2005 [2001]) that shows that you can modify the argument of Joyce (1998) to give the conclusion that degrees of belief ought to be Dempster-Shafer functions. Economics literature is rife with representation theorems for a vast array of different kinds of models of belief and utility. There are a number of weaker results in the mould of Fine's qualitative probability approach, for example Koopman (1940); Seidenfeld (1993); Seidenfeld, Schervish, and Kadane (1995). Whether Cox's theorem is modifiable in this way is, I think, an open question but I would guess that it is not so amenable to modification.

3.5.3. Are the arguments for probabilism mutually reinforcing?

Given the different motivations the various arguments rely on, and given that the arguments are successful to varying degrees on a number of points, it seems that there isn't much common ground where these arguments might be mutually reinforcing. If you are the kind of person who thinks that it's reasonable to restrict ourselves to finitely many propositions, then it looks like representation theorems and Cox's theorem are going to be unsatisfying. Then depending on whether you thought belief was tied more strongly to practicality or to epistemology, only one of the remaining two arguments would be appealing. If, on the other hand, you thought it was better that beliefs cover a rich algebra of propositions, then at least gradational accuracy would be ruled out. Say you thought it important that probabilism relate to conditional probability measures. This might rule out Dutch book arguments. If you thought degrees of belief were about your attitude to propositions, then representation theorems wouldn't seem to be doing the right thing at all. So wherever you stand, only one theorem is really convincing. In brief, given a particular attitude to what beliefs are, most arguments for probabilism are unconvincing.

One might criticise this line of thought by saying "but I care about practical rationality *and* epistemic rationality!" You might continue: "I think epistemic rationality suggests (via gradational accuracy) that beliefs ought to be probabilities, and practical rationality (via Dutch books) secures the same conclusion." The two arguments seem reinforcing despite their different starting points. But this misses the point. Dutch books secure the claim that degrees of belief *as action-guiding mental states* – as determinants of betting quotients – should be probabilities. Gradational accuracy says that degrees of belief *as estimates of truth values* should be probabilities. The burden of proof is on whoever claims these two ways of thinking of belief are the same. At the very least one needs to claim that estimates of truth values are a good guide to action. I'm not suggesting that such an argument can't be made, indeed, I think true beliefs will often be practically useful. But I do want to stress that the link between the two kinds of degrees of belief needs to be quite strong for the arguments to be mutually reinforcing. Arguably, Gibbard (2007) offers exactly the needed connection between practical rationality and epistemic rationality.

I think it is perfectly reasonable to say that my beliefs fail to satisfy the axioms of qualitative probability (for example, maybe my $\succeq_{\mathbf{b}}$ ordering is only partial), but I nevertheless consider the constraints on betting imposed by the Dutch book argument to be rational.[48] So while it is the case that I should act "as if" I had some probability function, subjectively, my beliefs don't have the structure of probabilities. The probabilist might respond that my phenomenological "beliefs" aren't doing any work, they aren't guiding action, so they aren't what's at stake. But that's just to deny that probability plays the role that the proponent of qualitative belief claims it does.

To make the claim that these arguments for probabilism are mutually reinforcing, one needs an argument to the effect that the different kinds of beliefs under discussion are tightly linked. That's not to say I don't think such an argument can be made, but just to point out that arguments for probabilism do not reinforce each other without such an argument. Let's say you wanted Joyce's nonpragmatic argument and a Dutch book type argument to be mutually reinforcing. You would need to argue that the sort of epistemic beliefs that Joyce is interested in and the practical action guiding beliefs under consideration in Dutch book arguments should have the same structure. And how would you go about arguing this, without being able to appeal to arguments for probabilism? Gibbard (2007) argues that a concern for accurate beliefs can only come from a desire for beliefs

[48]Something like this sort of idea is motivating the "Transferable Belief Model" which allows nonprobabilistic beliefs, but forces them to collapse into some probability function when you have to make a decision (Smets and Kennes 1994).

with high guidance value. And guidance value is the sort of practical thing that action-guiding beliefs should do well at.

Consider the following – admittedly rather odd – scenario. A crazy neuroscientist induces you to undergo an advanced brain scan that can ascertain whether or not you believe certain propositions. The neuroscientist tells you that if you have a certain false belief, he will reward you handsomely. Here it seems that value for the truth and pragmatic value have come apart: epistemic value is maximised by disbelieving the falsehood, while practical value is maximised by believing the falsehood. The reason is because of the strange way that the reward is tied directly to your beliefs rather than to your actions as informed by your beliefs.[49] Proponents of Gibbard's view needn't be moved much by such outlandish examples, but they do, I think, show that there is at least a conceptual wedge to be driven between practical and epistemic utility. What I take the above "crazy neuroscientist" example to show is that there's some implicit assumption about how beliefs influence value in Gibbard's argument.

It is certainly true that it is often practically valuable to have high credence in the truth, but that's a far cry from claiming that structurally, these disparate types of belief are similar. What about the claim that there is just one kind of thing – belief – and it fulfils all these roles. But that is just stating, by fiat, that the question under discussion has been decided. Belief is a slippery concept and I can't see any *a priori* reason to rule out pluralism. Another argument that you might offer for the claim that epistemic beliefs and action-guiding beliefs ought to be structurally similar is that there are independent arguments that each of these ought to have *this particular structure* and that therefore, they ought to have the same structure. But that is just putting the cart before the horse. An appeal could be made to theoretical unity at this point. We don't have strong reasons to believe that these disparate things we call "belief" have a common structure, nevertheless, there is some value in having a unified framework for modelling all of them, and probabilism provides that. This argument shouldn't, however, be taken to be an argument against exploring alternative frameworks for modelling belief when probabilism is found to be inadequate.

It does, however, suggest a different line of argument. Perhaps the very independence of the different arguments is a way in which they are mutually reinforcing. "See:" says the probabilist, "no matter what you think about the structure of

[49] This paragraph was in part inspired by a talk Hilary Greaves gave at a "Philosophy of Probability" conference in Oxford on "Cognitive Decision Theory".

the algebra of events, the conclusion is the same: probabilism. Whatever you think about the nature of belief, I have an argument that says you should be a probabilist." I don't think this is all that convincing. If I think beliefs should be action guiding in the way Dutch books assume they are, then why should Joyce's argument convince me at all? Given some position on the various considerations discussed above, it seems that some arguments are going to be irrelevant and therefore not convincing.

Maybe you could say "I have no firm position on what beliefs are, but whatever they are, there's an argument that says they're probabilities." This is a sort of "supervaluation" move. On all precisifications of the idea of belief, it is true that there is an argument for probabilism. So probabilism is "super-true". But this is only as strong as the weakest argument for probabilism. While this may give you a weak sort of "robustness" argument for probabilism, it doesn't mean that gradational accuracy arguments make Dutch book arguments more convincing, for example.

3.6. Probabilism as a regulative ideal

I have offered a range of arguments for probabilism, and I hope I have shown how each is flawed. However, I don't want to just completely ignore these arguments, and I don't want probabilism to have no role in my theory of uncertain belief. The criticisms I made of the various arguments' premises have been motivated in part by worries about severe uncertainty. So, if there is no severe uncertainty then probabilism should be warranted. Probabilism is serving as an ideal case, like the frictionless plane in physics. We are interested, in Henry Kyburg's nice phrase, in the "laws of frictionless thought" (Kyburg 1983, p. 232). Just as in the case of the frictionless plane, an idealisation can be an important part of getting to the heart of some particular phenomenon.

When you are certain of everything, your probability function reduces to a valuation function. In the same way, when you are sufficiently sure of things that there is no severe uncertainty, then your belief representor should reduce to a probability function. In a way, I think the move from probability to imprecise probability mirrors the move from valuation function to probability: the previous framework proves to be inadequate given the nature of the evidence (or lack thereof) that you have.

4. Imprecise probabilism

> On two occasions I have been asked, "Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?" I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.
>
> *(Charles Babbage)*

The previous chapter argued that the orthodox method for representing graded belief has some flaws. The main aim of this chapter is to outline my preferred alternative representation of belief. I motivate and defend the *imprecise probabilities* approach. More generally, this chapter further develops the kind of limited view of rationality that I subscribe to.

This choice, of course, encapsulates some particular value judgements about what a representation of belief is supposed to do. Not least, my view that only a weak standard of rationality is in play in cases of severe uncertainty. So I shall discuss that as well. As we saw in section 3.5, there are many reasons to think about formal representations of belief.

The plan is to first give some details of my preferred outlook on imprecise probabilism. Then I discuss how the theory should be understood. I then discuss a variety of ways to motivate the theory. After this, I briefly survey a couple of similar systems. Finally, I discuss learning and updating with imprecise probabilities.

4.1. What is imprecise probabilism?

The basic insight of the imprecise probabilities approach is to represent belief by a *set* of probabilities instead of a single probability. So instead of having some

pr represent your belief, you have $\mathcal{P}$, a set of such functions. Call this set your *representor*.[1] I will discuss various ways you might interpret the representor later (in section 4.2) but for now we can think of it as follows. Your representor is a *credal committee*: each probability function in it represents the opinions of one member of a committee that, collectively, represents your beliefs.

Often, it is assumed that your degree of belief in a proposition, X, is given by $\mathcal{P}(X) = \{\mathbf{pr}(X) : \mathbf{pr} \in \mathcal{P}\}$. I will adopt this notational convention, with the proviso that I don't take $\mathcal{P}(X)$ to be an adequate representation of your degree of belief in X. I return to this point later. From this we can derive an "imprecise expectation". If $\mathrm{E}_{\mathbf{pr}}(f)$ is the expected value of act f with respect to probability **pr**, then $\mathcal{E}_{\mathcal{P}}(f)$ is:

Definition 4.1.1 $\mathcal{E}_{\mathcal{P}}(f) = \{\mathrm{E}_{\mathbf{pr}}(f) : \mathbf{pr} \in \mathcal{P}\}$

The same proviso holds of $\mathcal{E}_{\mathcal{P}}(f)$ as held of $\mathcal{P}(X)$. I will often drop the subscript "$\mathcal{P}$" when no ambiguity arises from doing so.

From these concepts we can define some "summary statistics" that are often used in discussions of imprecise probabilities. Your "lower envelope" of X is: $\underline{\mathcal{P}}(X) = \inf \mathcal{P}(X)$. Likewise, your "upper envelope" is $\overline{\mathcal{P}}(X) = \sup \mathcal{P}(X)$. In the same way we can talk about the bounds of the spread of your expectation. Your "lower prevision" is defined as $\underline{\mathcal{E}}(f) = \inf \mathcal{E}(f)$ and your upper prevision ($\overline{\mathcal{E}}$) defined in the obvious way.[2]

One might want to call "lower prevision" "lower expectation", but we must be careful: Walley (1991) and Cozman (n.d.) use "lower expectation" to mean something slightly different. The difference won't be relevant here, so I stick with the more intuitive terminology of "expectation". Note however, that this is a (minor) deviation from what is standard in the literature. One might also prefer to term $\underline{\mathcal{P}}(X)$ a "lower probability" rather than a "lower envelope". I will allow myself to do this, but I should note that this is again slightly non-standard. Remember also that these summary statistics are not properly representative of your belief. Information is missing from the picture. I will have more to say about this later.

One might wonder whether taking representors to be sets of *conditional probabilities* as the basic item might not be a better way to do things. Hájek (2003) argues that conditional probabilities should be taken to be the basic entity, and

[1] This term is due to van Fraassen (1990).

[2] It hardly needs spelling out that if I am sceptical of $\mathcal{P}(X)$ as a representation of belief, I am doubly so of $\underline{\mathcal{P}}(X)$.

unconditional probabilities should be defined out of them. I agree with Hájek but in what follows, this point won't really make any difference, so I will continue to talk mostly of unconditional probabilities. Bear in mind that these are "really" conditional probabilities conditioned on $\top$.[3]

4.2. Interpreting the framework

How are we to understand claims like $\underline{\mathcal{P}}(X) = 0.3$? More generally, what interpretation should be given to the set of probabilities? Before we can ask this question, we need to ask an even more basic question.

4.2.1. What is a model of belief a model of?

Imprecise probabilities aren't a radically new theory. I see them as merely a slight modification of existing decision theory for situations of ambiguity. Often your credences will be precise enough, and your available actions will be such that you act more or less *as if* you were a strict Bayesian. I see imprecise probabilities as the "Theory of Relativity" to the strict Bayesian "Newtonian Mechanics": all but indistinguishable in all but the most extreme situations.[4] Howson (2012) makes a similar analogy between modelling belief and models in science. Both involve some requirement to be somewhat faithful to the target system, but in each case faithfulness must be weighed up against various theoretical virtues like simplicity, computational tractability and so on. Likewise Hosni (ms.) argues that what model of belief is appropriate is somewhat dependent on context. There is of course an important disanalogy in that models of belief are supposed to be *normative* as well as descriptive, whereas models in science typically only have to play a descriptive role.[5]

One standard interpretation of the probability calculus is that probabilities represent "degrees of belief" or "credences". This is more or less the concept that I have been considering so far. But what is a degree of belief? There are a number of ways of cashing out what it is that a representation of degree of belief is actually representing. It is important to get these concepts clear before we continue. I am

[3]Recall the discussion of section 2.3.2.

[4]This analogy goes deeper: in both cases, the theories are "empirically indistinguishable" in normal circumstances, but they both differ radically in some conceptual respects. Namely, the role of absolute space in Newtonian mechanics/GR; how to model ignorance in the strict/imprecise probabilist case.

[5]"Descriptive" is here being used in a fairly broad sense to include activities like prediction.

going to adopt the convention of using *degree of belief* to refer to the thing being modelled, and *credence* as the thing doing the modelling. I will try to keep to a similar convention as regards *value* (the thing modelled) and *utility* (the thing modelling).

One of the most straightforward understandings of degree of belief is that credences are interpreted in terms of an agent's limiting willingness to bet. Williamson (2010) takes this line. The idea is that your credence in X is q just in case q is the value at which you are indifferent between the bets (X, q) and $(\neg X, 1 - q)$. This is the "betting interpretation". This is the interpretation behind Dutch book arguments: this interpretation of belief makes the link between betting quotients and belief strong enough to sanction the Dutch book theorem's claim that *beliefs* must be probabilistic.

A related interpretation of credence is to understand credence as being just a representation of an agent's dispositions to act. This interpretation sees credence as that function such that your elicited preferences and actions can be represented as those of an expected utility maximiser with respect to that function. Your credences just are that function that represents you as a rational agent.[6] A slightly more sophisticated version of this sort of idea is to understand credence to be exactly that component of the preference structure that the probability function represents in the representation theorem.

One might take the view that credence is modelling some kind of mental or psychological quantity in the head. Strength of belief is a real psychological quantity and it is this that credence should measure. Unlike the above views, this interpretation of credence isn't easy to operationalise. It also seems like this understanding of strength of belief distances credence from its role in understanding decision making. The above *behaviourist* views take belief's role in decision making to be central to or even definitional of what belief is. This psychological interpretation seems to divorce belief from decision. Whether there are such stable neurological structures is also a matter of some controversy.

A compromise between the behaviourist views and the psychological views is to say that belief is characterised *in part* by its role in decision making. This leaves room for belief to play an important role in other things, like assertion or reasoning and inference. So the answer to the question "What is belief?" is: "Belief is whatever psychological factors play the role imputed to belief in decision making

[6]For precise probabilism, "rational agent" means "expected utility maximiser". For imprecise probabilism, rational agent must mean something slightly different.

contexts, assertion behaviour, reasoning and inference". There is room in this characterisation to understand credence as measuring some sort of psychological quantity that causally relates to action, assertion and so on. This is a sort of functionalist reading of what belief is. Eriksson and Hájek (2007) argue that "degree of belief" should just be taken as a primitive concept in epistemology. The above attempts to characterise degree of belief then fill in the picture of the role degree of belief plays.

I want to be a little non-standard in my understanding of belief. I don't want to focus solely on human-sized agents and their personal beliefs. I'd also like to use the same formal framework to theorise about the explicit decision weights used in real cost-benefit analyses. These don't necessarily match any particular human agent's personal beliefs: they are what might be called "institutional credences". These are the sort of credences of interest in scientific decision making, which will be the topic of a later chapter.[7]

In any model of uncertain belief and uncertain inference, there are some things about which you are uncertain. But in every case there are some things that are taken for granted. Imagine a toy example of drawing marbles from an urn. The observed frequencies of colours is used as evidence to infer something about the frequencies of colours in the urn. In this model, you take for granted that you are accurately recognising the colours and are not being deceived by an evil demon or anything like that. That's not to say that we couldn't model a situation where there was some doubt about the observations: the point is that in the simple case, that sort of funny business is just ruled out. There are certain aspects of the situation that are taken for granted: that are outside the modelling process. This is the same in science: when we model the load stress of a particular design of bridge, we take for granted the basic physics of rigid bodies and the facts about the materials involved. This point relates to what I was saying earlier about models of belief being like models in science. It also relates to the discussion of Lo and Mueller (2010) and the idea of level of uncertainty being relative to what you take for granted (section 1.1).

[7]Of course, in this case I am moving away from having belief be tied to *psychological* quantities. But belief is still somehow tied up with whatever structural features of the institutional set up and the beliefs of the individuals involved that are involved in decision making.

4.2.2. What makes belief rational?

Once we've decided what sort of thing belief is, we need to decide what we will mean by *rational* belief. There are levels of normativity here. The most basic constraints on rational belief, those that almost everyone will assent to, are basic structural constraints: it is unreasonable, irrational to believe contradictory things. For precise probabilists these structural constraints will include adherence to the axioms of probability theory. Imprecise probabilists will not accept this structural constraint in the same form, although probability theory still maintains some of its normative force for them. Some might maintain that the basic structural constraints are all that is required for rationality. I call these people *radical subjectivists*.

One might also demand that rational beliefs should be responsive to the evidence, in a certain sense. Williamson (2010) calls this "empirically-based Bayesianism". Note here we are talking about a synchronic norm: your current evidence should inform your current credences. There's a separate issue of how new evidence should change your credences, but I don't want to discuss that yet.[8] If you thought that obeying the structural constraints and being in accord with the current evidence was all there was to rationality, and that as long as you did that, you could believe what you like, I would term you an *empirical voluntarist*.

W.K. Clifford argued that

> ...it is wrong always, everywhere and for anyone to believe anything on insufficient evidence Clifford (1901, p. 175)

This suggests that there needs to be a procedure for determining what the correct belief should be when the evidence underdetermines belief. Clifford seems to be thinking in terms of outright (full) belief and so the correct attitude would be suspension of judgement. If you are a Bayesian, you need to do a little work to find out what the appropriate level of belief is. The *Objective Bayesian* has one such method of determining the correct belief when the evidence is weak. I find myself sympathetic to objectivists but, as I discuss later, I am sceptical of the details of Objective Bayesianism.

William James criticised Clifford's objectivism and argued for a kind of *voluntarism* that I mentioned above (James 1897). James starts by agreeing with the intuition motivating Clifford's remarks:

[8] Indeed, radical subjectivists will argue that adherence to a diachronic evidential norm is enough to guarantee that credences are suitably responsive to evidence.

> All this strikes one as healthy, even when expressed by Clifford with somewhat too much robustious pathos in the voice. Free-will and simple wishing do seem, in the matter of our credences, to be only fifth wheels to the coach. (p. 8)

He goes on to point out that there seem to be competing epistemological pressures.

> Believe truth! Shun Error! — These, we see, are two materially different laws; and by choosing between them we may end by coloring differently our whole intellectual life. (p. 18)

James worries that Clifford is too focused on shunning error and that any such commitment will lead to an impractical level of scepticism. The widespread suspension of belief that James thinks Clifford's view leads to is, he claims, impossible. When the evidence doesn't decide what to believe for us, we must believe *something*, and nothing but free choice is left to decide what that belief should be. I take it to be important to any sort of objectivist project in the imprecise probabilities mould to respond to this worry about the danger of a looming scepticism.

So there are three levels of normativity for rational belief. One might consider only structural constraints. One might think that empirical constraints limit your choice of belief. Or one might think that empirical constraints *determine* your choice of belief.

In the kinds of scientific decision making contexts that I will eventually return to, these are the kind of probabilities of interest. In this context, I feel, objectivism certainly has the upper hand over any form of permissivism or voluntarism. Whatever might be the case of real individual agents' freedom over their beliefs, institutional credences shouldn't be based on hunches, gut feeling or what have you: only the evidence should count. Given that we are understanding these institutional credences as explicit decision weights used in decision making, there must be a "paper trail". At some stage, the use of *these* rather than *those* weights might have to be justified: reasons given for why the cost benefit analysis sanctioned *this* rather than *that* course of action. Jamesian relativism precludes having the *reasons for belief* that are necessary in this context.

Even at the level of desiring that your evidence determine your beliefs, I think there is a certain level of subjectivity or arbitrariness that is unavoidable. First, there is choice of a formal language to describe the objects of belief. This choice could have an impact. It certainly has an impact for Objective Bayesianism, as Williamson (2010) points out in chapter 9. It is perhaps a little less clear whether this language-relativity is as big a problem for the imprecise case. Whether or not it is, there are other elements of subjectivity that enter the process. When setting up the model of belief, I have to decide what things to take for granted – what facts to take as part of the background information. There is also the question of ascertaining exactly what import various kinds of evidence you might receive should have. To take a simple example, what should the import be of observing 66 heads out of 99 tosses of a coin? One would like to say, at the very least, that this rules out the possibility of the coin's being heavily biased toward tails. But it is logically possible that a tails-biased coin could come up heads 66 out of 99 tosses.[9] The question is "What standard of evidence do we need to assent to some claim?" The standard of logical proof is obviously too strong. Indeed, it seems like logic alone cannot determine what standard of evidence we require. There is some subjectivity in fixing the standards for evidence: but once this is done, the evidence determines your belief. In different ways, Kyburg's evidential probabilities model (Kyburg 1983; Kyburg and Teng 2001) and the unreliable probabilities model of Gärdenfors and Sahlin (1982) both build in this kind of subjectivity. Kyburg by the thresholds for acceptance, Gärdenfors and Sahlin with their reliability threshold ρ_0. Both of these models will be discussed a little later. I will return to this point later, but for now I want to point out that the ideal of a *purely objective* model of belief isn't going to work. What I mean by this is that it seems like principles of rationality alone don't seem to determine what language you should frame your beliefs in, nor do they determine what standard of evidence is appropriate. That's not to say that these things are *subjective* – they may just be inescapable aspects of your psychology – but just that they are not determined by the principles of rationality alone. We can and should mitigate and constrain how the elements of subjectivity enter the process.

4.2.3. What represents belief?

So now we have a better idea of what it is that a model of belief should do. But which part of our model of belief is representing which part of the belief state? The first thing to say is that $\mathcal{P}(X)$ is *not* an adequate representation of the belief in X. That is, one of the values of the sets of probabilities approach is to capture certain kinds of non-logical relationships between propositions that are lost when focusing on, say the associated lower probability. For example, consider tossing

[9]The extreme finite frequentist might disagree: that it comes up heads more often tails *makes it the case* that it isn't tails-biased, so what I described is impossible.

a coin of unknown bias. $\mathcal{P}(H) = \mathcal{P}(T) = [0,1]$, but this fails to represent the important fact that $\mathbf{pr}(H) = 1 - \mathbf{pr}(T)$ for all $\mathbf{pr} \in \mathcal{P}$. This complementarity can play an important role in reasoning and decision. This is a point I will return to several times. I mentioned it when I introduced "$\mathcal{P}(X)$", I am mentioning it again now, and I will return to this theme when I discuss updating later (section 4.6).

$\mathcal{P}(X)$ might be a *good enough* representation of belief for some purposes. For example in the Ellsberg game which we shall meet later (section 4.3.3), these sets of probability values (and their associated sets of expectations) are enough to rationalise the paradoxical preferences. How good the representation needs to be depends on what it will be used for. Representing the sun as a point mass is a good enough representation for basic orbital calculations, but obviously inadequate if you are studying coronal mass ejections, solar flares or other phenomena that depend on details of the internal dynamics of the sun.

4.2.4. Imprecise and indeterminate

Sometimes a distinction is made between *imprecise* credences and *indeterminate* credences. The idea is that there are two distinct kinds of belief state that might be modelled by sets of probabilities. An imprecise belief is one where there is some probability function that reflects your preferences, but you don't know what it is, so you use a set. An indeterminate belief is a belief state where there is genuine indeterminacy in your beliefs. Steele (2007) argues that it is the indeterminate beliefs that are interesting.

I am following Walley (1991) in using the term "imprecise probability" to cover cases of indeterminate probability. This is an unfortunate aspect of a somewhat standard terminology. It has been suggested to me that "indeterminate probabilities" is a better name for the items I am interested in. I am somewhat inclined to agree, but this advice came far too late in the day for me to effect such a terminological upheaval in the current project. In any event, "imprecise probability" is still infinitely better than the term "mushy credence".

4.2.5. Interpretations of the representor

There are a number of ways to understand what $\mathcal{P}$ is supposed to represent. I mention a couple here, although a better handle on the interpretation is probably gained from looking at the arguments in favour of the position, which I shall turn to in a moment.

One way one might want to interpret $\mathcal{P}$ is as a set of plausible chance functions, for a given confidence interval. One could interpret Kyburg's Evidential Probabilities in this way[10] (see section 4.5.1).

The worry with viewing $\mathcal{P}$ as a set of possible chance functions is that it doesn't seem to be properly responsive to evidence. Imagine tossing a biased coin, with chance of heads either $\frac{1}{3}$ or $\frac{2}{3}$. So if this were the right interpretation of the representor, $\mathcal{P}(H) = \left\{\frac{1}{3}, \frac{2}{3}\right\}$. Since coin tosses are independent, evidence of previous coin tosses would not change anything about the representor. Intuitively however, seeing 66 out of 99 tosses land heads is good evidence that the true chance is $\frac{2}{3}$. See Halpern (2003, Example 3.4.1, pp. 84–5).

The above worry – that sets of chances aren't responsive to new evidence suggests another interpretation. This interpretation of the representor that is due to Joyce (2011)[11] is to consider each probability in the representor as a member of a *credal committee*. This committee represents all the possible prior probabilities you could have that are consistent with the evidence. Each credal committee member is a fully opinionated Jamesian voluntarist. The committee as a whole, collectively, is a Cliffordian objectivist.

The idea is that the elements of your representor are ways that sufficiently better evidence might have determined your credence. There is something somewhat "supervaluationist" about this understanding of the representor. Indeed, connections between imprecision and vagueness are deep and interesting. I shall discuss them again later. One can think of the committee as containing members whose beliefs encode all the possible situations that need to be taken into account. Collectively, the committee must reason and decide. The overall committee opinion somehow supervenes on the opinions of the members of the committee. Each committee member, whatever the prior belief, is sensitive to the evidence. Thus, representors so interpreted *should* respond to evidence appropriately. Later we will see that this is not quite right.

Finally, one might take imprecise probabilities to just be a representation of "fuzzy" or "incomplete" attitudes. This would motivate having credal states represented by intervals or sets of numbers rather than precise values. The sets of probabilities approach might then be seen as a solid formal foundation for that sort of representation.

[10] Though this isn't the only interpretation of Evidential Probabilities.

[11] Joyce attributes it Adam Elga. Weatherson (ms.) discusses something similar.

4.3. Conceptual arguments for imprecise probabilism

In the last chapter I spent a long time discussing various arguments for precise probabilism that have been put forward. I ultimately find the arguments wanting. Not because they are deeply flawed, but because they impose too strong constraints on your belief. Given the kinds of severe uncertainty I am concerned with, probabilism seems like an unattainable ideal. In this section I want to offer some positive arguments why the imprecise framework is better than its precise cousin.

I offer some epistemological, conceptual arguments; and, in the next section, some arguments of a more formal character.

Most of the following epistemological or conceptual arguments in favour of imprecise probabilism can be seen as different attempted articulations of a common idea. When modelling belief, it is almost universally agreed that it is important to reflect the evidence you have in the belief you hold. Especially in the case where you are modelling belief for the purpose of decision making, evidence should be represented. I claim that there is an important corollary of this principle: you ought to represent your ignorance. That is, it is important to have your representation of belief reflect your evidence, but it should also reflect where your evidence is lacking. In a number of ways, precise probabilism glosses over this important point.

4.3.1. Incomparability and indifference

In chapter 3 I discussed some arguments that made use of preference relations, or qualitative belief relations. In all these cases, the relation was assumed to be complete.[12] In cases of severe uncertainty, I don't consider completeness to be reasonable. That is – to take the preference example – I think it reasonable to have no preference in either direction. This is, I feel, an importantly different attitude to being *indifferent* between the options (see section 3.2.7).

Is it really reasonable to say that an agent has a preference between any two possible acts?[13] In the context of eliciting an agent's preference, it might be a

[12] In section 3.1 the preference was only assumed to have a strictly weaker property, but what I have to say here holds of this case too.

[13] Indeed, given the Rectangular Field Assumption and the Richness axiom, the space of acts over which you are expected to have a preference is unreasonably vast.

reasonable thing to assume, since "having neither preference" is more or less behaviourally indistinguishable from "being indifferent" (i.e. having both preferences). But I think there is an important difference between positively holding both weak preferences and having neither preference. In the context of the kinds of uncertainty I am interested in, I think this difference will be important, so lumping "no preference" together with "indifference" is not justified. As I have said above, I am coming from a quite different direction from those motivated by behaviourist concerns. My concern is to represent your epistemic/doxastic state so as to facilitate good decision making in cases of severe uncertainty. A representation theorem is besides the point.

The question I am asking is "given the current state of evidence, what is the right belief state to have in order to best make the right decisions?" I take it that it is possible for evidence to not decide which of two hypotheses is more likely, which of two possible actions will do better. I think it is important that our model of belief should accommodate these facts.

Various of the arguments for probabilism required unreasonable things of you. Representation theorems required you to have preferences over astonishingly rich spaces of acts; Dutch book theorems required you to be able to determine preference between arbitrary bets: would you rather bet on getting at least one 1 on four rolls of a die or on getting at least one pair of 1s on 24 rolls of two dice?[14] In either case it seems like there are discriminations of preference that are required of you that you simply cannot make.[15] From this and the contrapositive of an "ought-implies-can" principle, we conclude that the conjunction of the axioms cannot be normative. More specifically, at least some of the axioms are not normative constraints. Arguably completeness and richness are among those axioms that shouldn't be considered normative; Joyce (1999) agrees. We can now ask what can be proved in the absence of the purely structural axioms? This surely gives us a handle on what is really required of the structure of belief. The specific answer to this question will be left until the section on formal arguments (section 4.4).

[14] This is apparently the problem that bankrupted the Chevalier de Méré and caused him to turn to Blaise Pascal for an explanation. This in turn led to the birth of probability theory.

[15] Perhaps this example won't convince if you take on board the idea that our agents are logically omniscient. But just imagine bets on propositions where you know nothing about the nature of the propositions. Surely here is a case where indifference is not warranted. That is, replace the lack mathematical ability with some evidential failing in the example.

4.3.2. Weight of evidence, balance of evidence

Joyce (2005) suggest that there is an important difference between the *weight* of evidence and the *balance* of evidence. This is a distinction that collapses in precise probabilism. Joyce argues – and I agree – that the distinction is worth representing. This idea has been hinted at by a great many thinkers including J.M. Keynes, Rudolf Carnap, C.S. Pierce and Karl Popper (See references in Gärdenfors and Sahlin 1982; Joyce 2005). Imprecise probabilism allows you to represent this difference.

Consider tossing a coin known to be fair. Let's say you have seen the outcome of a hundred tosses and roughly half have come up heads. Your degree of belief that the coin will land heads should be around a half. This is a case where there is weight of evidence behind the belief.

Now consider another case: a coin of unknown bias is to be tossed. That is, you have not seen any data on previous tosses.[16] In the absence of any relevant information about the bias, symmetry concerns might suggest you take the chance of heads to be around a half. This opinion is different from the above one. There is no weight of evidence, but there is nothing to suggest that your attitudes to H and T should be different. So, on balance, you should have the same belief in both. However, when offered the choice between betting on the fair coin's landing heads as opposed to the unknown-bias coin's landing heads, it doesn't seem unreasonable to prefer the former. But if both coins have the same subjective probabilities attached, what rationalises this preference?[17] So these two different cases get represented as having the same probabilistic belief. Joyce argues that there is a difference between these beliefs that is worth representing. Imprecise probabilism *does* represent the difference. The first case is represented by $\mathcal{P}(H) = \{0.5\}$, while the second is captured by $\mathcal{P}(H) = [0, 1]$. Another way to understand what is wrong with the precise probabilist's response to this coin example is due to Karl Popper (quoted in Gärdenfors and Sahlin (1982, p. 369)). If the precise probabilist must have $\mathbf{pr}(H) = 0.5$ before any evidence has arrived, and if the evidence (E) that arrives is that the coin lands heads about half of the time, then the posterior belief will be $\mathbf{pr}(H|E) = 0.5$. The evidence, therefore is irrelevant: it has had no impact

[16]The phrase "coin of unknown bias" is a little awkward, since there's no sense in which what is unknown is a *bias*. Bias with respect to what? The coin could only said to be biased if we had a pre-existing idea that the coin *should* land tails as often as heads. We should perhaps speak of a "coin with unknown chance of heads". Or if chance-talk is not allowed, "coin with disposition to generate unknown statistics". Hykel Hosni was helpful in pressing me on this point.

[17]This is related to the Ellsberg cases I mention in a moment (section 4.3.3).

on the belief.

Kaplan (1996, 2010) makes the same point. He suggests that to conflate the two cases discussed above is to commit the "sin of false precision" (Kaplan 1996, p 23–32). He notes that it is not the case that it is always unreasonable or impossible for you to have precise beliefs: in that case precision could serve as a regulative ideal. Kaplan's argument is that to expect to always have precise beliefs is to fail to give evidence its due (p. 29). I think precise probabilism does still serve as something of a regulative ideal, but it is the belief of an ideal agent *in an idealised evidential state*. Idealised evidential states will include things like tossing a coin known to be fair.[18]

Scott Sturgeon perhaps puts the point best when he says:

> *[E]vidence and attitude aptly based on it must match in character*. When evidence is essentially sharp, it warrants sharp or exact attitude; when evidence is essentially fuzzy – as it is most of the time – it warrants at best a fuzzy attitude. In a phrase: evidential precision begets attitudinal precision; and evidential imprecision begets attitudinal imprecision. Sturgeon (2008, p. 159, Sturgeon's emphasis)

Hawthorne (2005) argues that Bayesians need both degree-of-belief and also degree-of-support measures. I think this is pushing in the same direction as Joyce's argument. Hawthorne's motivation and his solution are different, but ultimately, both his project and the current argument are trading on the idea of degree of belief and degree of support coming apart, and both being important aspects of belief. The basic idea in both cases is that it's not just what the evidence says, but also *how much* evidence there is.

It might be objected that the difference between weight and balance of evidence can be captured within a single probability measure by attending to the extent to which updating on various kinds of evidence would change the posterior belief. In the case where there has already been some evidence, extra evidence won't do much to shift the belief in Heads. However, in the case of no evidence, the belief will be more prone to shifts due to new evidence. Brian Skyrms identifies a property that plays something of this role, but in the context of propensities and lawlikeness. He calls it "Resiliency" (Skyrms 2011).

If all we wanted was *something* that corresponded to the distinction between weight and balance of evidence, then this would be fine. However, if we want the

[18]Or, more properly, a coin *taken* to be fair.

distinction to be of relevance in *decision making*, then it needs to be represented in such a way that it can impact on decisions. The imprecise probabilities approach can do this because the difference in belief is relevant to decision. Arguably, the single measure approach cannot. Or rather, the standard decision theory is not sensitive to this difference in the possible future updated beliefs. We could of course come up with some sort of "resiliency-weighted expected utility" which would nullify this objection. And presumably at least some of the other objections to standard probability theory can be dealt with the same way. But the more we do this, the epicycles on epicycles we add to our precise probabilist framework, the less clear it becomes that this has any advantage in simplicity over the imprecise framework.

Strictly speaking, what I have argued here is only half the story. I have shown that there *is* a difference between the representations of belief as regards weight and balance. But that still leaves open the question of exactly what is representing the weight of evidence? We could, of course, just measure the amount of evidence in each case. The fair coin case is like the unknown-bias case, but you have some extra information. In this sense, the beliefs in the former are based on stronger weight of evidence. But what aspect of the belief reflects this? It seems unsatisfactory to simply use $\overline{\mathcal{P}}(H) - \underline{\mathcal{P}}(H)$ as a measure of the weight of evidence for H. This would get wrong cases of conflicting evidence. (Imagine two equally reliable witnesses: one tells you the coin is biased towards heads, the other says the bias is towards tails.) We can, I think, sidestep this problem. The advantage of imprecise probabilities in these cases is not that they represent the weight of evidence *per se*, but that they allow a representation of *the difference between weight and balance* such that we can rationalise differences in behaviour that we think are due to this distinction.

Sadly, it is not always true that imprecise probabilities represent this difference. In cases of *dilation* (see section 4.6.2) more evidence can cause the upper and lower probabilities to get further apart. I defer an analysis of this awkward fact until I have presented the phenomenon of dilation in detail.

4.3.3. Ellsberg problems

There are a number of examples of decision problems where we are intuitively drawn to violate precise probabilism. And indeed, experimental subjects do seem to express preferences that violate the axioms. Imprecise probabilities offers a way of representing these intuitively plausible and experimentally observed choices as rational. One classic example of this is the *Ellsberg paradox*. Here I adapt a version of the problem from Halpern (2003).

> Example 1 I have an urn that contains a hundred marbles. Thirty marbles are red. The remainder are blue or yellow in some unknown proportion.

What are we to think about this example? The standard precise probabilist beliefs would have $\mathbf{pr}(R) = 0.3$, and then, using some symmetry reasoning, we get $\mathbf{pr}(B) = \mathbf{pr}(Y) = 0.7/2 = 0.35$.

These probabilities suggest that you should prefer a bet on yellow to a bet (for the same stakes) on red. In the betting formalism from section 3.1, you should have the following preference: $(B, 0.3) \succeq (R, 0.3)$. However, experimental evidence – for example Camerer and Weber (1992) – suggests that real agents often prefer bets on red to bets on yellow and also prefer red to blue. These revealed preferences are difficult to square with the standard precise probabilist story. That is, no probability function determines an agent who has the following preferences:

- $(B, 0) \sim (Y, 0)$
- $(R, 0) \succeq (B, 0)$
- $(R, 0) \succeq (Y, 0)$
- $(R, 0.3) \sim (B \vee Y, 0.7)$

The first three of these reflect the experimentally observed (and intuitively reasonable) preferences, the fourth reflects the intuition that red should have a betting quotient of 0.3 reflecting its known probability.

The imprecise probabilist can model the situation as follows: $\mathcal{P}(R) = 0.3, \mathcal{P}(B) = \mathcal{P}(Y) = [0, 0.7]$.[19] Modelling the ambiguity allows us to rationalise real agents' preferences for bets on red. To flesh this story out would require a lot more to be said about decision making. I defer this task until the next chapter, but the intuition is that aversion to ambiguity explains the preference for red over blue or over yellow.

As Steele (2007) points out, the above analysis rationalises the Ellsberg choices only if we are dealing with genuinely indeterminate beliefs. If we were dealing with a case of "unknown precise belief" then there would exist some **pr** in the

[19]Note that this expression of the belief state misses out some important details. For example, for all $\mathbf{pr} \in \mathcal{P}$, we have $\mathbf{pr}(B) = 0.7 - \mathbf{pr}(Y)$. For the point I am making here, this detail is not important.

representor such that rational choices maximise $E_{\mathbf{pr}}$. For the Ellsberg choices, there is no such **pr**.

Let me be clear here that I take the current project to be a normative one. So what is important here is the ability of the imprecise probability model to rationalise the intuitively compelling response to this decision problem. Even a normative theory should, perhaps, have one eye on its descriptive adequacy so that the intuitively reasonable results are also the experimentally observed is reassuring. But the important feature is that the imprecise probability model can capture some intuitively reasonable behaviour that the standard account cannot. Put it this way: the intuitive Ellsberg preferences put a constraint on what the norms can say. The norms should rationalise behaviour that we consider reasonable. Thus, descriptive adequacy in this case points to success as a normative theory. One might argue the other way: one might argue that the norms of precise probabilism are rationally compelling and that therefore the observed behaviour is irrational. One man's modus ponens is another man's modus tollens.

Here is another decision problem. This one is set up to make clear that the agents are violating an apparently plausible principle of rationality known as the Sure Thing Principle. I will discuss this principle in more detail in chapter 5.

	R	B	Y
$L_1 = (R, 0)$	1	0	0
$L_2 = (B, 0)$	0	1	0
$L_3 = (R \vee Y, 0)$	1	0	1
$L_4 = (B \vee Y, 0)$	0	1	1

Table 4.1.: Another decision problem

The bets involved in this problem can be shown as in Table 4.1 (using the same urn set up as before). The paradoxical preferences that people tend to exhibit are $L_1 \succeq L_2$ but $L_4 \succeq L_3$. These pairs of bets differ only in what happens in the "Y" column, so the preference should go the same way in each case. That is, L_1 and L_3 look the same if you ignore the Y column. Likewise for L_2 and L_4. Now, in the choice between L_1 and L_2, the Y column shouldn't make a difference, since you get the same thing if Y from both bets. Likewise for L_3 and L_4. So you should choose the same way in both choices. That is, $L_1 \succeq L_2$ iff $L_3 \succeq L_4$. These are not the preferences people tend to express.

Again, a similar analysis to the one above allows the imprecise probabilist to rationalise these preferences.

4.3.4. Objective Bayesianism without precision

I find myself sympathetic to much of what Williamson (2010) argues in defence of Objective Bayesianism.[20] Williamson offers a list of criteria for an adequate theory of rational belief. I agree with the spirit of Williamson's norms, but not the precise forms he appeals to. Williamson's norms are: Probabilism, Calibration and Equivocation. I will say a little about each in turn. It turns out that I agree with the spirit of each norm, but I disagree with the exact versions Williamson advocates.

The first of the norms of Objective Bayesianism is Probabilism. This is the claim that degrees of belief should be probabilities. I feel I have probably already said enough to make clear what my opinion of this norm is. Williamson uses what he calls the "betting interpretation" of degrees of belief. That is, he takes your having a degree of belief p in X to *mean* that p would be the betting quotient you would agree to buy or sell bets on X. This set up should be familiar from section 3.1. But the use it is put to is different. Williamson is serious about interpreting degrees of belief as simply meaning having a particular disposition to accept certain betting quotients as fair.

Williamson accepts, in effect, what I earlier called Complementarity. He is aware of the apparent unreasonableness of this as a constraint on betting preference, but argues that it is appropriate for an interpretation of degrees of belief. That is, he concedes that actual models of real-world betting are better off not taking Complementarity as a constraint, but that this does not preclude the "fair buy/sell betting quotient" being a reasonable explication of strength of belief. And given this, the Dutch Book Argument secures the conclusion that degrees of belief must be probabilistic.

As I spent some time discussing in the previous chapter, I don't buy this conclusion. I do however accept that probability theory serves as a regulative ideal: in the absence of severe uncertainty, strength of belief *is* appropriately modelled by probability theory.[21] But when there is severe uncertainty – when Complementar-

[20] With the proviso that I don't quite buy the centrality of pragmatic "avoiding sure loss" arguments.

[21] I actually think epistemic utility arguments are stronger than the pragmatic betting related arguments for probabilism. Williamson doesn't consider these kinds of arguments.

ity is unwarranted – the DBA sanctions DS belief functions. So if I were to accept a modified version of Williamson's "betting interpretation of belief", then the appropriate model of belief would be Dempster-Shafer theory. But I don't agree with this.[22] I take one part of the understanding of degree of belief to be its role in decision making. But this role is not exhausted by the setting of betting quotients. So while the buying quotients and selling quotients do somehow constrain belief, DS theory isn't the complete theory of belief.

The second norm of Objective Bayesianism is Calibration. The spirit of the Calibration norm is that your degrees of belief should be sensitive to your evidence. As Hume puts it:

> A wise man... proportions his belief to the evidence.
>
> Hume (1975 [1777], p. 110)

Williamson considers a variety of formal ways of cashing out what this sensitivity to evidence should amount to (pp. 39–49). All of these various ways essentially put constraints on what probability functions would count as beliefs consistent with the evidence. So an evidential constraint is essentially a set of probability functions, and Calibration is the claim that your belief should be in the set of probability functions determined by your evidential constraints. Williamson considers two broad kinds of evidence: evidence of chances and structural constraints. The first kind of evidence consists of constraints like $\{\mathbf{pr}(X) = x\}$ or $\{\mathbf{pr}(X) \in [x, y]\}$. These are essentially constraints that tell you what the chance of a certain proposition is. Principles like Miller's principle or Lewis' Principal Principle are examples of precursors of the Calibration norm. In our Ellsberg example, we had a constraint of this form: $\{\mathbf{pr}(R) = 0.3\}$.

The second sort of evidential constraint – the structural constraints – will be things like "X and Y are independent"[23], or "X is more likely than Y". In the Ellsberg example the constraint $\{\mathbf{pr}(B) = 0.7 - \mathbf{pr}(Y)\}$ is a constraint of this form.

Williamson offers a pragmatic argument for the chance constraint (pp. 40–2), but doesn't offer much in the way of justification of the structural constraints. In any case Calibration seems justifiable on directly epistemological grounds. *Of course* your degrees of belief should be responsive to your evidence! That's what they are for. How exactly the evidence you have imposes constraints on the permissible set of probabilities is tricky. Statistical evidence might inform constraints

[22] Later, in section 4.5.3, I discuss my reasons for not being content with DS theory.

[23] There are, in fact, many distinct notions of independence for sets of probabilities. See Cozman (2012).

of the first kind. Evidence of causes might inform structural constraints.

The exact constraints that evidence puts on belief are difficult to pin down, as Williamson agrees. But I certainly endorse the idea that evidence should constrain belief. What I would like to highlight here is that Williamson uses sets of probabilities to represent the constraints that evidence imposes of belief.

The constraints that evidence puts on belief can't be *mere* consistency. That is, it can't be that a belief is ruled out only when it is formally inconsistent with the evidence. Consider the evidence that 666 of 999 tosses of a coin of unknown bias have landed heads. It is *consistent* with this evidence that the coin is heavily biased towards tails. However, intuition suggests that in this case the evidence should determine that belief in heads should be around $\frac{2}{3}$. Wheeler and Williamson (2011) offer the beginnings of an understanding of how the Calibration norm should be cashed out using Henry Kyburg's work on Evidential Probabilities. Deciding what beliefs should be considered consistent with the evidence will involve an element of subjectivity. I discuss Evidential Probabilities in a little more detail in section 4.5.1.

The final norm of Objective Bayesianism is Equivocation. This norm says that degrees of belief should be as equivocal as is consistent with the evidence. That is, Calibration determines a set of possible probability functions, and Equivocation determines which of those should be used as your belief: it picks the most equivocal one. The most equivocal member of $\mathcal{P}$ is denoted $\Downarrow\mathcal{P}$. This is a necessary step since Williamson's version of the Probability norm only allows for a single probability to represent belief. Equivocation can be seen as one way of cashing out what I think of as the "dual norm" of Calibration: you should not believe anything more strongly than is sanctioned by the evidence. I agree with this idea: that evidence constrains belief, and that going beyond your evidence is also unreasonable. However, I take issue with Williamson's cashing out of this idea in terms of his Equivocation norm. I will say two things about this. First, I will discuss Williamson's argument from caution. Second, I will discuss Williamson's argument against the interval-valued probabilist.

Equivocation, for the Objective Bayesian, amounts to choosing the probability function consistent with the evidence that is closest to the equivocator function. This is a special probability function that assigns the same probability to all atoms of the algebra of events. The argument for this Equivocation norm is another pragmatic argument: equivocal beliefs lead to more cautious behaviour and thus minimise worst-case expected loss. Williamson's argument relies on the

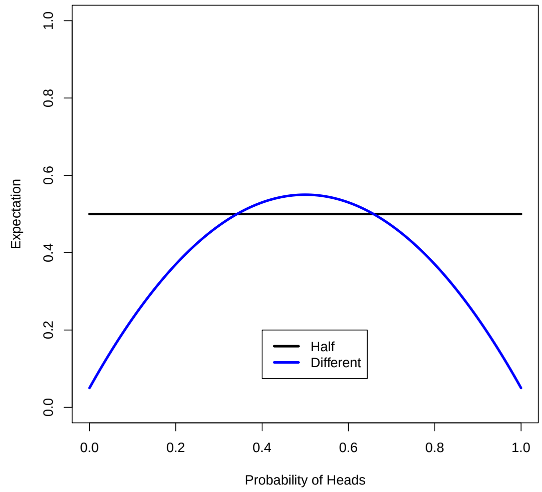

Figure 4.1.: The decision problem of Example 2

following claim: "the risks associated with actions triggered by middling degrees of belief tend to be less severe than those triggered by extreme degrees of belief." (Williamson 2010, pp. 57–8). The argument for this is an appeal to intuitions: you can afford to take risky decisions only when you are confident of the outcome (that is, only when you have extreme beliefs). But I think Williamson is conflating degree of belief and degree of support. The following example serves to illustrate the difference.

Example 2 You are betting on a coin of unknown bias. It will be tossed twice. The following bets are available to you.

Half Win £0.50

Different Win £1 + ε if the coin lands on different sides each time, win ε otherwise

	HH	HT	TH	TT
Half	0.5	0.5	0.5	0.5
Different	ε	$1+\varepsilon$	$1+\varepsilon$	ε

As is clear from Figure 4.1, Different maximises expectation only when $\mathbf{pr}(H)$ is very close to a half. Imagine you have no evidence about the bias of the coin. It seems the high risk action – the action with higher variance in expectation – is the one sanctioned by a middling credence. Extreme beliefs in this case would sanction the constant (zero risk) action.[24] If you were now given lots of evidence that the coin lands heads very nearly half the time, then and only then does it seem reasonable to choose betting on Different. This illustrates that it isn't high degree of belief that triggers high-risk actions, but it is high degree of support. The two are indistinguishable at the level of beliefs in the Objective Bayesian framework; this prompts Williamson to conflate them.

So I don't think Williamson's argument from caution works. Williamson is committed to getting close to the equivocator being the right way to have non-committal beliefs. I would argue that taking the full set of probabilities consistent with the evidence is the more non-committal approach. Williamson meets this challenge head-on (pp. 68–72).[25] There are two arguments against the imprecise view that I wish to address.

First, Williamson argues that imprecise probabilities can fail to determine an optimal action. He may be right about this, although a more sophisticated analysis of imprecise decision theory would be required to properly ground this claim (see chapter 5). But even if it is true that imprecise probabilities make decision making hard, this is not a bug: it's a feature! In the presence of severe uncertainty *decision making should be hard*. It's obvious that if you have very little evidence, then it should be difficult to determine what the best course of action is. So Williamson's claim that Objective Bayesianism is better because it makes decision theory easier is misguided: it makes decision theory *too easy*. There are only some circumstances when "Newtonian mechanics is easy, so we should use it rather than GR" is a good argument. Important, decision relevant uncertainties are lost when you equivocate. It is helpful to know when your evidence doesn't make choice among the options straightforward. This should focus the decision maker on other aspects of the available options, like their reversibility, their robustness, their adaptability and so on. This line of argument will take centre stage in a later chapter. This argument can be recast in terms of my weaker version of rationality. Williamson argues that imprecise probabilities might fail to determine a best course of action. I argue that such a failure is an advantage because imprecise probabilities fail to determine an optimal action just when we shouldn't expect rational constraints on choice to determine such an action.

Williamson's second argument against imprecise probabilities is a direct re-

[24]The figure has $\varepsilon = 0.05$. If the figure isn't sufficiently convincing, consider that as ε tends to 0, Different tends to a weakly dominated act.

[25]Williamson is actually concerned with interval-valued probabilities, but his comments carry over to the full imprecise probabilities framework.

sponse to the claim that Objective Bayesianism fails to distinguish weight of evidence and balance of evidence. Williamson argues that on the contrary, imprecise probabilism is wrong to make the distinction. This is because weight/balance is a distinction concerned with *evidence*, not with *belief*. The claim is that the distinction should not be represented at the level of belief. Williamson is right that the weight/balance distinction is one that is primarily about evidence. But, I would argue that the weight/balance distinction is decision relevant. And if it is belief – not evidence – that informs decision,[26] the distinction needs representing on the level of belief. If I am right about Williamson's conflating of degree of belief and degree of support in his discussion of extreme beliefs leading to risky actions, then he has already really conceded that both parts of the distinction are decision relevant.

Ultimately, my objection boils down to this: I feel it is important that the full range of the set of probabilities be represented. The equivocation norm is the wrong way to be true to Clifford's maxim. We shouldn't want to try to reduce our uncertainty down into a single probability measure. Representing the diversity of the calibrated probability functions is important.

4.3.5. Suspending judgement

A precise Bayesian can't even ask when it is reasonable to suspend judgement. That is, the precise probabilist does not have the resources to represent the state of withholding belief. This is related to the distinction between weight and balance mentioned above.

Consider the coin of unknown bias again. The Bayesian agent must have a precise belief about the coin's landing heads on the next toss. Given the complete lack of information about the coin, it seems like it would be better just to suspend judgement. That is, it would be better not to have any particular precise credence. But there just isn't room in the Bayesian framework to do this. The function must output some number.

It is helpful to look at the relationship between DS theory and probability theory in this context. One interpretation that is given of the Dempster-Shafer belief function $\mathbf{bel}(X)$ is that it represents the degree to which the available evidence supports X. Then $\mathbf{bel}(\neg X)$ is the degree to which the evidence supports $\neg X$. One can then consider the quantity $I(X) = 1 - \mathbf{bel}(X) - \mathbf{bel}(\neg X)$. This measures the degree to which the evidence is silent on X. Huber (2009) points out that precise

[26]Williamson tacitly concedes this in using his model of belief to discuss decisions.

probabilism – which is a special case of DS theory – can then be understood as making the claim that no evidence is ever silent on any proposition. That is, $I(X) = 0$ for all X. One can never suspend judgement. This is a nice way of seeing the strangeness of the precise probabilist's attitude to evidence.

The committed precise probabilist would respond that setting $\mathbf{pr}(X) = 0.5$ *is* suspending judgement. This *is* the maximally noncommittal credence. I would argue that this only *seems* to be the right way to be noncommittal if you are wedded to the precise probabilist approach. Suspending judgement is something you do when the evidence doesn't determine your credence. But for the precise probabilist, there is no way to signal the difference between suspension of judgement and strong evidence of probability half. This is just the weight/balance argument again.

To make things more stark, consider the following delightfully odd example from Adam Elga:

> A stranger approaches you on the street and starts pulling out objects from a bag. The first three objects he pulls out are a regular-sized tube of toothpaste, a live jellyfish, and a travel-sized tube of toothpaste. To what degree should you believe that the next object he pulls out will be another tube of toothpaste? Elga (2010, p. 1)

In this case, unlike in the coin case, it really isn't clear what intuition says about what would be the "correct" precise probabilist suspension of judgement. What Objective Bayesian Equivocation recommends will depend on seemingly arbitrary choices about the formal language used to model the situation.[27]

Another response to this argument would be to take William James' response to W.K. Clifford. James argued that as long as your beliefs are consistent with the evidence, then you are free to believe what you like. So there is *no need* to ever suspend judgement. Thus, the precise probabilist's inability to do so is no real flaw. This attitude, which I called *epistemic voluntarism* has some adherents, but also some detractors. The main line of argument against voluntarism is that there is no *reason* for your belief's being thus and so, rather than some other way.

I side with the Objective Bayesian in thinking that it is better for your beliefs to not have this subjective, unreasoned element. Your evidence should not just constrain your belief; it should determine it. There is then a need to be able to suspend judgement when evidence is silent. That is, we are interested in *rational*

[27] Williamson is well-aware of this language relativity problem. He argues that choice of a language encodes some of our evidence.

belief: belief informed by rational constraints. When the evidence is silent, it seems that rationality cannot but require that you suspend judgement.

On the subject of suspending judgement, consider the question of logical omniscience again. In the case of *full* or *categorical* belief, you can be logically ignorant in two distinct ways: you can believe the negation of a consequence of your other beliefs; or you can simply fail to believe some consequence of your beliefs. The second is clearly a lesser sin, and the modifications to your beliefs required in order to accommodate believing the consequence are easier to achieve.[28] In the graded belief case, there is no such distinction: a precise probabilist must have some particular value attach to every proposition. There is only one way to be logically ignorant; which is to fail to believe consequences of a proposition more strongly than the proposition itself. There is no easy way to effect the appropriate belief change in order to accommodate learning of your mistake for the precise probabilist. I take this to show that if we wanted to explore logical ignorance and graded belief, an imprecise probabilist framework, with its resources for suspending judgement, would be better suited than a precise probabilist framework.

4.3.6. Regress, vagueness and practicality

Imprecise probabilities is a theory born of our limitations as reasoning agents, and of limitations in our evidence base. If only we had better evidence, a single probability function would do. But since our evidence is weak, we must use a set. In a way, the same is true of precise probabilism. If only we knew the truth, we could represent belief with a valuation function, or just a set of sentences that are fully believed. But since there are truths we don't know, we must use a probability. And indeed, the same problem arises for the imprecise probabilist. Is it reasonable to assume that we know what set of probabilities best represents the evidence? Perhaps we should have a set of sets of probabilities... Similar problems arise for theories of vagueness (Sorensen 2012). We objected to precise values for degrees of belief, so why be content with sets-valued beliefs with precise boundaries? This is the problem of "higher-order vagueness" recast as a problem for imprecise probabilism. Why is sets of probabilities the right level to stop the regress at? Why not sets of sets? Why not second-order probabilities? Why not single probability

[28]In the belief revision literature this can be made precise: Remedying the situation where you fail to believe a consequence is achieved by an addition. Remedying the situation where you believe the negation of a consequence is achieved by a revision, which is (by the Levi identity) a contraction then an addition (Hansson 2011).

functions? This is something of a pragmatic choice. The further we allow this regress to continue, the harder it is to deal with these belief-representing objects. So let's not go further than we need.

I have argued above that imprecise probabilism does have some advantage over precise probabilism, in the capacity to represent suspending judgement, the difference between weight and balance of evidence and so on. So we must go *at least* this far. But for the sake of practicality, we need not go any further. This is, ultimately, a pragmatic argument. Actual human belief states are probably immensely complicated neurological patterns with all the attendant complexity, interactivity, reflexivity and vagueness. We are *modelling* belief, so it is about choosing a model at the right level of complexity. If you are working out the trajectory of a cannonball on earth, you can safely ignore the gravitational influence of the moon on the cannonball. Likewise, there will be contexts where simple models of belief are appropriate: perhaps your belief state is just a set of sentences of a language, or perhaps just a single probability function. If, however, you are modelling the tides, then the gravitational influence of the moon needs to be involved: the model needs to be more complex. I am arguing that an adequate model of belief under severe uncertainty needs to move beyond the single probability paradigm. But a pragmatic argument says that we should only move as far as we need to. So while you need to model the moon to get the tides right, you can get away without having Venus in your model. This relates to the contextual nature of appropriateness for models of belief that I mentioned earlier.

4.3.7. Nonprobabilistic chance

What if the objective chances were not probabilities? If we endorse some kind of connection between known objective chances and belief – for example, Lewis' Principal Principle (Lewis 1986) or a form of calibration norm – then we might have an additional reason to endorse imprecise probabilism. It seems to be a truth universally acknowledged that chances ought to be probabilities, but it is a "truth" for which very little argument has been offered. For example, Schaffer (2007) makes obeying the probability axioms one of the things required in order to play the "chance role", but offers no argument that this should be the case. Joyce (2009) says "some have held objective chances are not probabilities. This seems unlikely, but explaining why would take us too far afield." (p. 279, fn. 17). Various other discussions of chance – for example in statistical mechanics (Frigg 2008b; Loewer 2001, 2004) or "Humean chance" (Lewis 1986, 1994) – take for granted that chances should be probabilistic. Obviously things are confused by the use of the concept of chance as a way of interpreting probability theory. There is, however, a perfectly good pre-theoretic notion of chance: this is what probability theory was originally invented to reason about, after all. This pre-theoretic chance still seems like the sort of thing that we should apportion our belief to, in some sense. And there is very little argument that chances must always be probabilities. If the chances were nonprobabilistic in a particular way, one might argue that your credences ought to be nonprobabilistic in the same way.

I want to give a couple of examples of this idea. First consider some physical process that doesn't have a limiting frequency but has a frequency that varies, always staying within some interval. It might be that the best description of such a system is to just put bounds on its relative frequency. If we took a Humean perspective on what chances are, this would make it the case that its chance is nonprobabilistic (Hájek and Smithson 2012). If you were committed to apportioning your credence to the chances, then in worlds with these sorts of entities, you ought to have nonprobabilistic credences. If instead you lived in a world of genuine indeterminacy, then imprecise statistics would also move you towards nonprobabilistic credence (I discuss imprecise statistics in the next section).

4.4. Formal arguments for imprecise probabilism

I have above discussed some conceptual, epistemological reasons to consider imprecise probabilism superior. I now consider more formal arguments. These include some subversions of the arguments for probabilism that occurred in the last chapter, as well as some new ones.

4.4.1. Completions of incomplete preferences

As should be clear from the previous discussion, I am no fan of complete preferences (section 3.2.6,3.2.7). However, I don't believe that complete preferences are irrational: I simply feel that completeness of preference is a counsel of perfection. Since there is an Ought-Can principle at work, and since you cannot always have complete preferences because of severe uncertainty, it follows that we must not demand that you ought to have such preferences. My objection to completeness is a "not-ought" objection, not an "ought-not" objection.

Precise probabilism is serving as a regulative ideal. Completeness of preference is what the thoroughly informed agent ought to have.[29] Without complete preference, Savage's representation theorem doesn't work. However, for each *completion* of the incomplete preference ordering, the theorem follows. So if we consider the set of probability functions that are such that some completion of the incomplete preference is represented by that function, then we can consider this set to be representing the beliefs associated with the incomplete preference. We also get, for each completion, a utility function unique up to linear transformation.

A similar kind of argument can be given for the richness assumptions. That is, the uncountable infinity of states required to ensure the uniqueness of the probability function was considered unreasonable in the previous chapter. If this richness is considered, not as a condition on reasonable models of decision problems, but as a condition of *extendibility*, then it isn't as unreasonable. The idea is that your state space may be finite or countably infinite, but it should at least in principle be possible to extend it consistently to a richer one. So again, consider each possible sufficiently rich extension of a finite or countable state space. For each extension, the representation theorem holds. Taking the set of probability functions thus acquired delivers us an imprecise probability representation.

So, sets of probabilities offer a representation of incomplete preferences or incomplete qualitative belief orderings. If all we wanted was a way to represent incomplete beliefs, what need would we have of this framework, however? Surely the preference ordering or the belief ordering itself would serve as a good enough representation. However, there are kinds of belief state, kinds of opinion, that I feel I have and that the ordinal representation fails to adequately represent. For example, I can think "X is *much* more likely than Y", or that "X and Y are independent[30]". These kinds of belief attitude aren't captured by the ordinal representation. So the set of probability approach allows one to represent these more subtle kinds of belief state. Thus we should move beyond the merely ordinal representation of belief.

Given that a particular purely ordinal preference ordering that satisfies Savage's postulates will have a unique probability representation, it could be argued that the "is twice as likely" type information is *somehow* in the ranking. The point, however, is that it isn't represented in a way that we can easily see it. If we want a

[29] I don't wish to wade into the debate about genuinely incommensurable goods. For the purposes of the current debate, let's assume there are no such basic incommensurabilities in play. But see Hsieh (2008).

[30] Obviously, I can't be appealing to probabilistic independence here, but there is an intuitive sense in which events can be independent.

model where we can discuss this kind of "slightly more than ordinal" information, the brute preference ordering isn't appropriate, even if it is in some sense formally adequate to express those aspects of belief.

The purely ordinal approach to representing belief can't do justice to the action guiding dimension of decision theory. Consider the choice between bets (X,α) and (Y,β) where $X \succ_{\mathbf{b}} Y$ but $\beta \prec_u \alpha$. That is, X is more likely than Y, but the bet on Y leads to a better prize (the bet costs less). How are we to trade off the better prize versus the bet more likely to win? It depends on the strength of belief and *how much* the prizes are valued. The representation theorems take these sorts of preferences among bets – or lotteries or acts more generally – and uses them to determine cardinal measures of belief and value. But if you want to use decision theory as a way to help us determine these preferences by using your attitudes to the events and prizes, then the purely ordinal approach isn't sufficiently expressive to allow you to make all the determinations you might want.

4.4.2. Unknown correlations

Haenni et al. (2010) motivate imprecise probabilities by showing how they can arise from precise probability judgements.[31] That is, if you have a precise probability for X and a precise probability for Y, then you can put bounds on $\mathbf{pr}(X \wedge Y)$ and $\mathbf{pr}(X \vee Y)$, even if you don't know how X and Y are related. These bounds give you intervals of possible probability values for the compound events.

For example, you know that $\mathbf{pr}(X \wedge Y)$ is bounded above by $\mathbf{pr}(X)$ and by $\mathbf{pr}(Y)$ and thus by $\min\{\mathbf{pr}(X), \mathbf{pr}(Y)\}$. If $\mathbf{pr}(X) > 0.5$ and $\mathbf{pr}(Y) > 0.5$ then X and Y must overlap. So $\mathbf{pr}(X \wedge Y)$ is bounded below by $\mathbf{pr}(X) + \mathbf{pr}(Y) - 1$. But, by definition, $\mathbf{pr}(X \wedge Y)$ is also bounded below by 0. So we have the following result: if you know $\mathbf{pr}(X)$ and you know $\mathbf{pr}(Y)$, then, you know that $\mathbf{pr}(X \wedge Y)$ is somewhere in the interval:

$$[\max\{0, \mathbf{pr}(X) + \mathbf{pr}(Y) - 1\}\ ,\ \min\{\mathbf{pr}(X), \mathbf{pr}(Y)\}]$$

Likewise, bounds can be put on $\mathbf{pr}(X \vee Y)$. $\mathbf{pr}(X \vee Y)$ can't be bigger than when X and Y are disjoint, so it is bounded above by $\mathbf{pr}(X) + \mathbf{pr}(Y)$. It is also bounded above by 1, and thus by the minimum of those expressions. It is also bounded below by $\mathbf{pr}(X)$ and by $\mathbf{pr}(Y)$ and thus by their maximum. Putting this together,

[31] Apparently Seidenfeld and Levi also favour this motivation.

$\mathbf{pr}(X \vee Y)$ will be in the interval:

$$[\max\{\mathbf{pr}(X), \mathbf{pr}(Y)\}\ ,\ \min\{\mathbf{pr}(X) + \mathbf{pr}(Y), 1\}]$$

So if your evidence constrains your belief in X and in Y, but is silent on their interaction, then you will only be able to pin down these compound events to certain intervals. Any choice of a particular probability function will go beyond the evidence in assuming some particular correlation between X and Y. That is, $\mathbf{pr}(X)$ and $\mathbf{pr}(X|Y)$ will differ in a way that has no grounding in your evidence.

4.4.3. Imprecise statistics

The statistical evidence that we often use to constrain our beliefs does not, as a rule, give us precise probabilities. This is the motivating idea behind the work of Peter Walley (Walley 1991, 2000; Walley and Fine 1982). This is related to a similar idea propounded by Kyburg and Teng (2001), which I will discuss later.

A contrived example will serve to illustrate the idea behind what I mean by "imprecise statistics". Imagine you are collecting statistics on hair colour, and you have a number of categories: Grey, Blonde, Brown, Black... And imagine you are collecting data in the evening. You see someone who has dark hair, but you can't tell whether they have brown or black hair, given that the light is fading. You might be tempted just to guess one way or the other and put the subject in one category. However, the motivating idea behind imprecise statistics suggests putting the subject in the disjunctive category "Brown or Black" without putting them in either of the disjunct categories. The statistics thereby built up would be superadditive, but not additive.

4.4.4. Imprecise scoring rules

This argument has some of the same flavour as the "completions of incomplete preference" idea above. If epistemic utility arguments appeal to you, but you can't decide on exactly which inaccuracy measure to use, then take the set of all probability functions that accuracy dominate your nonprobabilistic belief and take this set as the right belief.

One might want to go further and try to generalise epistemic utility functions such that they can measure the utility of sets of measures. Then one might try to prove that sets of probability functions are vindicated. I don't know of any work in this area.

Perhaps it would help to step back and ask the more basic question. What makes a set of probabilities a good belief? First let's look at the question of what makes a good probabilistic belief. Say you make a prediction that there's a 66% chance of a "barbecue summer". That summer in fact turns out to be unusually cold. You take a lot of flak for having made a bad prediction. How aggrieved should you be? More generally, when is a probabilistic prediction a good one? There is quite a literature on this question. For example, see Bröcker and Smith (2007).

One can score beliefs in the same way you can score probabilistic forecasts. This gives you a way to answer the question "What is a good belief?". The same question arises in the imprecise probabilities setting. What is a good imprecise belief? Can scoring rules be modified to answer this question? One crude attempt at scoring imprecise beliefs would be to take the set of inaccuracy values for each $\mathbf{pr} \in \mathcal{P}$ as representing the inaccuracy of the representor. Note the analogy to taking sets of expectation pointwise. We could then consider the upper and lower limits of this set of inaccuracies. I don't know of any work on this, and am sceptical that this crude approach will yield much in the way of interesting results.

If the aim were to use some sort of scoring rules approach to vindicate imprecise probabilism, we would have to be careful to set things up so as to not give an advantage to singleton probabilities. For example, if the inaccuracy of the representor were its average inaccuracy,[32] then there would always be a precise probabilistic belief with lower inaccuracy: that's how averages work! Implicit in the DS approach to belief – with its understanding of "suspend judgement" to mean $\mathbf{bel}(X) = 0 = \mathbf{bel}(\neg X)$ – is the idea that it is better to underestimate a chance that to overestimate it. Presumably if we built scoring rules that built in this asymmetry then they would vindicate DS theory over probability theory. In fact, it is some sort of asymmetry of this sort that was at work in my earlier Dutch book theorem for DS theory: focus on the buying odds (the α_Xs) means that we end up with DS belief functions. Williams (2012a) points to an analogy between Dutch book arguments and gradational accuracy. Betting success could be seen as a kind of score, so this gives hope for the possibility of "asymmetric" gradational accuracy arguments to vindicate nearby theories.[33]

It is well known that the Brier score can be decomposed into parts, each of which can be given an intuitive interpretation (Joyce 2009, Section 12). The decomposition in fact holds of any strictly proper scoring rule (Bröcker 2009). The terms

[32] Let's not worry for now what I mean by average.

[33] The decision problems of Frigg et al. (2013b) set up an asymmetry by only offering selling odds, and not buying odds. So the conclusion there is effectively a DS **plaus** measure.

of the decomposition are a calibration term and a discrimination term. The first tells you how reliable your belief was and the second tells you how informative your belief was.[34] Presumably any sort of asymmetry in the scoring rule should be built into the calibration part, rather than the discrimination part.

4.4.5. Non-classical logic

We saw earlier that at least the Dutch book and gradational accuracy arguments can be subverted by replacing the underlying logic by a non-classical one. If you were a big fan of intuitionistic logics of some variety, for example, then perhaps these facts would convince you to abandon probability theory. However, this route would not lead you directly to a sets-of-probabilities approach, since the conclusions of the above mentioned arguments sanctioned Dempster-Shafer theory. Indeed, it is built into both frameworks that some sort of single-function model of belief will be chosen.

So to make this sort of argument work for *me*, I would need to argue that sets of probabilities offer a good explication of DS theory and other nearby theories. I will do a bit of this in section 4.5. But I don't consider this my main motivation for imprecise probabilities, so I won't discuss this too much.

4.5. Nearby theories

Imprecise probabilities do not exist in a vacuum. There are many similar alternatives to precise probabilism in the vicinity. I discuss a few of the more important and more interesting ones.

4.5.1. Evidential probabilities

Evidential probabilities are a kind of "interval-valued" probability together with a thorough theory for inferring them from statistical data, and for logically reasoning about them. The theory is most fully set out in Kyburg and Teng (2001) and more clearly in Haenni et al. (2010); Wheeler and Williamson (2011). The theory is discussed more with an eye to the concerns of psychologists and experimental economists in Kyburg (1983). The fundamental guiding principle behind

[34]This is following Joyce. The discrimination term can be further broken down into terms that Bröcker calls "uncertainty" and "resolution". The details aren't important at this level.

the inference from data step comes from "Reichenbach's principle", which can be explained as follows:

> If we are asked to find the probability holding for an individual future event, we must first incorporate the case in a suitable reference class. An individual thing or event may be incorporated in many reference classes... We then proceed by considering the narrowest reference class for which reliable statistics can be compiled.
>
> Reichenbach quoted in Kyburg and Teng (2001, p. 77)

Given a particular significance level, we take statistically significant sample frequencies as evidence of population frequencies. The significance level is used to determine the "Evidential Corpus" of evidence claims that determines belief. We give a kind of "confidence interval" of values for the population frequency. We take this interval as an interval-valued representation of the correct belief about the property in question. The significance level, which is somewhat subjective, can be thought of in this context as a sort of "threshold for reliability of statistics". Relative to this choice of significance level, the impact of evidence is objective and determinate. So evidential probabilities are relative to a choice of significance level. This is an element of subjectivity in the theory.

Wheeler and Williamson (2011) discuss using evidential probabilities as a method for putting restrictions on which probability functions are compatible with a certain corpus of statistical data. If we stop there – rather than going the further step and equivocating – then we can take evidential probabilities as sanctioning a particular (imprecise) representor as the correct belief to have about the world.

I don't take evidential probabilities to be a rival to my brand of imprecise probabilism, so much as a method for inferring representors from data. Indeed, I am sympathetic to Wheeler and Williamson's interpretation of evidential probabilities as providing a method for carrying out the Calibration procedure of Objective Bayesianism.

As typically understood, Evidential Probabilities are interval-valued functions, or at least set-valued functions. They are not sets of probability functions. However, it is easy enough to understand the interval valued function as determining a set of probabilities, $\mathcal{P}$, such that $\mathcal{P}(X)$ is exactly the interval for X for all X.

4.5.2. Unreliable probabilities

Gärdenfors and Sahlin (1982) introduce a theory that they term *Unreliable Probabilities*. It bears some resemblance to Kyburg's theory – they in fact discuss Kyburg – but it is a theory built with decision making in mind. The basic idea is that you have a set of probabilities and attached to each probability function is a measure of its reliability. Depending on the circumstances, you pick some reliability threshold ρ_0 and restrict your attention to the set of probabilities that are at least as reliable as that threshold. They then have a story about how decision making should go with this set. I defer discussion of the decision part to a later chapter. Note that they don't really need a *measure* of reliability, all they need is something to *order* the probabilities in $\mathcal{P}$. The threshold then becomes some cut-off probability: anything less reliable than it doesn't make the cut.

Gärdenfors and Sahlin don't really discuss this measure of reliability a great deal. Presumably reliability increases as evidence comes in that supports that probability function. Gärdenfors and Sahlin offer an example to illustrate how reliability is supposed to work. They consider three tennis matches. In match A, you know that the two tennis players are of roughly the same level and that it will be a tight match. In match B, you have never even heard of either player and so cannot judge whether or not they are well matched. This should remind you of the weight/balance distinction I discussed earlier. In match C you have heard that the players are really unevenly matched: one player is much better than the other, but you do not know which of the players is significantly better. If we graphed reliability of a probability against how likely that probability thinks it is that the player serving first will win,[35] the graphs would be as follows: graph A would be very sharply peaked about 0.5; graph B would be quite spread out; graph C would be a sort of "U" shape with high reliability at both ends, lower in the middle.

Graph A is peaked because you know that the match will be close. You have reliable information that the player serving has about a 50% chance of winning. Graph B is spread out because you have no such information in this case. In case C, you know that the probability functions that put the chances at near 50–50 are unreliable: all you know is that the match will be one-sided.[36]

The index on Kyburg's evidential corpora and the current reliability index

[35] Let's assume there are no other events in our algebra in order to guarantee that each probability value for server-wins corresponds to one and only one probability. This is for illustrative purposes only.

[36] Case C here is an excellent example of why convexity of imprecise probabilities should not be mandated.

threshold both indicate a place where an element of subjectivity enters rational belief. Where to set that cut-off affects which things get counted as beliefs and which things influence learning, deliberation, choice, action, assertion and so on. The threshold is somewhat arbitrary, perhaps context sensitive, but not rationally derivable in any case.

What might be part of deciding how reliable **pr** $\in \mathcal{P}$ is? One idea we might use is that probabilities that have scored well at predicting past evidence (by the lights of some scoring rule like the Brier score, for instance) are reliable. Probabilities that consistently assign high probability to events that don't happen and low probability to events that do happen will not score well and will thus be considered unreliable. This method of scoring the reliability of probabilities will help with some of the problems we encounter in section 4.6.

What might be deciding what the threshold for inclusion in $\mathcal{P}$ is? Earlier I suggested that it could include an element of subjectivity. It should vary with context. Perhaps the threshold should be something like "anything more than 90% of the most reliable **pr** gets in".

With this framework we can begin to ask some interesting questions: such as "what are the preconditions for reliability to covary with distance from the equivocator?" Williamson seems to think that reliability always covaries with distance from $\Downarrow \mathcal{P}$; I have discussed above why I think this is mistaken. However, the two things will often coincide and it would be interesting to see whether we can understand more carefully what the relationship is.

4.5.3. Dempster-Shafer theory

We have already met some of the formal details of Dempster-Shafer theory in section 2.3.2. Here I outline a little more broadly the motivations behind the project.

The DS belief function $\mathbf{bel}(X)$ is considered to represent the degree to which the evidence supports X. To return to the imprecise statistics example, we can see that the evidence that the subject has dark hair supports the proposition that the subject has brown or black hair, but one might take the view that the evidence isn't sufficient to support either disjunct.[37]

It is well known that DS belief functions are intimately related to imprecise

[37] I'm not suggesting that this interpretation of the situation is uncontroversial: a committed precise Bayesian might argue that dark hair is equally strong evidence for each disjunct, or something along those lines.

probabilities. Every DS belief function can be associated with a set of probabilities such that the lower probability of that representor is that DS function.[38] This is the content of Halpern's theorem 2.4.1 (Halpern 2003, p. 34).

The problem, I feel, with DS theory is that it makes decision theory harder, without really providing the advantages of imprecise probabilities. But I want to just focus on one way that decision making with DS belief functions seems to get things wrong. Let's imagine that we want DS decision making to work more or less like it does for probabilities. That is, we want to take the **bel** function and just drop it in as a replacement for **pr** throughout our decision making discourse. So agents should be acting to maximise:

$$\sum_i \mathbf{bel}(X_i)u(a(X_i)) \tag{4.1}$$

There's a little subtlety here. What is that i ranging over? That is, what are the X_i? Is it ranging over the atoms? Is it ranging over some partition such that each act is constant on each element of the partition? Since probabilities are additive, this doesn't matter at all for the precise probabilist. But for DS belief, it matters a lot! Consider this pair of bets on a coin of unknown bias: $\{(H, 0.4), (T, 0.4)\}$. This should straightforwardly come out as being an acceptable set of bets whatever your beliefs. But consider the DS-expected utility on the partition $\{H, T\}$:

$$\mathbf{bel}(H)(1-0.8) + \mathbf{bel}(T)(1-0.8) \tag{4.2}$$

Now, if **bel** is such that $\mathbf{bel}(H) = 0 = \mathbf{bel}(T)$, as would be reasonable for complete ignorance of the coin's bias, then this set of bets is valued at zero. Valuing this set of bets at 0 determines preferences that are in conflict with Bet Dominance. However, the act of taking both bets also happens to be constant on the trivial partition $\{\top\}$ since $H \vee T \equiv \top$. With this partition, the DS-expected value comes out as 0.2.

There are two lessons to take from this. First, DS-expected utility seems to be sensitive to choice of partition. I take this to be a flaw of a decision theory. This sort of difference of presentation should not affect the choiceworthiness of the acts. Second, some choices of partition seem to lead to undesirable choice profiles. The natural way to pick a partition in a principled way would be to just sum the **bel**-weighted utilities over the atoms. But this choice is most likely to lead to odd results, as equation (4.2) shows. I suppose you could try to argue that the correct partition is the coarsest partition such that all the acts are constant on all

[38]The set of probabilities is not unique. There is, however, a unique *convex* set.

the elements of that partition. This would rule in favour of the trivial partition in the example. But this sort of rule looks ad hoc, and vulnerable to odd results through gerrymandering the set of bets on offer to manipulate the partition.

Imprecise probabilities don't have this problem. Expectation is done pointwise for each member of the representor, and since each $\mathbf{pr} \in \mathcal{P}$ is a probability, the choice of partition doesn't matter. These problems for DS theory are also problems for many similar theories: Choquet expected utility, upper and lower probabilities, interval valued probabilities and so on. In each case, the way out of the problem is to find the imprecise probability that best matches the representation, and use that as the basis for decision making. As I suggested above in the case of Evidential Probability, one can interpret an interval valued probability as the set of probability functions with values in the intervals.

4.6. Updating beliefs

So far, I have only discussed "static" or "synchronic" structure on belief. But beliefs change in response to evidence. I won't have too much to say about this in the following chapters, but I feel *something* should be said about updating sets of probabilities. There are two reasons for this. First, no formal model of belief would be complete without a discussion of how belief should respond to new evidence. Second, imprecise probabilities suffer from some problems with updating.

4.6.1. Kinds of learning

There are, as I see it, several kinds of updating. I shall describe a few here. I don't mean to suggest that this list is exhaustive, nor that each kind of learning requires its own mechanism. But there do seem to be interesting distinctions to draw between each of the following kinds of learning.[39]

Learning a proposition

The simplest kind of learning is arguably learning that a proposition is true. This is the sort of learning that is straightforwardly modelled by conditionalisation

[39] For an alternative characterisation of types of learning, see Bradley (2009) where three types of learning are considered – revision, formation and withdrawal – with parallels to the literature on AGM-type revision for full belief.

in the precise case. The most obvious way to update sets of probabilities in this case is to take sets of updated probabilities. And it seems natural to understand "updated probabilities" as meaning conditionalised probabilities. Grove and Halpern (1998) offer a theorem that says that given certain conditions on how updating should work, sets of conditionalised probabilities is the only method for updating imprecise probabilities. Their conditions are not all intuitive and I don't mean to discuss them further.

An argument for a slightly weaker conclusion goes as follows. Whatever else happens, updating for singleton sets of probabilities should be exactly like updating single probability functions. This is effectively a kind of "boundary condition" on updating. We want things to reduce to conditionalisation in the precise limit. The imprecise probabilities framework I am developing is supposed to be a somewhat conservative generalisation of orthodox Bayesianism. Take $\mathcal{P}_E$ to be $\mathcal{P}$ updated on evidence E. This first condition amounts to:

$$\text{If } \mathcal{P} = \{\mathbf{pr}\} \text{ then } \mathcal{P}_E = \{\mathbf{pr}(\cdot|E)\} \tag{4.3}$$

Second, if $\mathcal{P}$ contains all the functions in $\mathcal{P}'$, then updating $\mathcal{P}$ should result in a representor that contains all of the updated version of $\mathcal{P}'$. That is:

$$\text{If } \mathcal{P}' \subseteq \mathcal{P} \text{ then } \mathcal{P}'_E \subseteq \mathcal{P}_E \tag{4.4}$$

This second condition can be seen as a commitment to a kind of "pointwise" view of how updating should work. It denies that any kind of holistic properties of the representor should impact on how each committee member updates. Together, these imply that an updated representor should contain at least all the conditional probabilities of its members.[40] The procedure that takes the updated representor to be exactly the set of conditional probabilities has been called *generalised conditioning*.

A related, more restricted kind of updating has been contemplated by Gilboa and Schmeidler (1993). They take the updated representor to consist of only those conditionalised probabilities that assigned the evidence a high prior probability. They call this "Maximum likelihood" updating. This is a "subset" of generalised conditioning, in that the ML update is always contained in the GC update.

Learning a constraint

Imagine if you were to learn "X is more likely than Y". This seems like it is interestingly different from learning the truth of a proposition. This sort of

[40]Proof: $\mathcal{P} \supseteq \mathbf{pr}$ for all $\mathbf{pr} \in \mathcal{P}$. So $\mathcal{P}_E \supseteq \{\mathbf{pr}(\cdot|E)$ for all $\mathbf{pr} \in \mathcal{P}$.

updating might best be captured by recalibrating. That is, by repeating the calibration procedure we discussed in section 4.2.2 and section 4.3.4.

Henry Kyburg's Evidential Probabilities aren't updated by conditioning. At each stage, the same process is conducted: given the current state of evidence, the evidential probability of X is the proportion of Xs in the smallest reference class for which we have good statistics. When new evidence comes in, the EP in X could change if the evidence determines statistics for a smaller reference class containing X.

The Objective Bayesian paradigm also rejects conditionalising as a method of updating. Like Evidential Probabilities, OB uses the same procedure to determine credence at each stage. Evidence determines the set of consistent probabilities, and then you must equivocate over them. If we ignore the equivocation part, we have a view of learning where new evidence can change the belief in X if the evidence rules out some previously consistent probability functions.

The difficulty that these sorts of approaches have is to do with diachronic consistency and reasoning about the future. It seems reasonable that what you imagine your beliefs would be were you to learn E and what your beliefs actually are once you learn E should match. In orthodox Bayesianism, the conditional probability plays a double role in being both belief under supposition of E and belief updated by E. So responses to evidence that take this recalibration route must tell a convincing story about how supposing-that-E beliefs relate to learning-that-E beliefs. If supposing that E involves doing the updating "in a subjunctive mood", or something like that, then there will be no inconsistency, but it will probably be computationally intensive. The nice thing about the orthodox Bayesian way is that the supposing-that-E beliefs are *already* part of your credence.

Undermining evidence

Another possible sort of learning is to discover that some source of evidence previously thought to be reliable turns out not to be. You then have to reassess all the evidence that came to you from that source in the light of its unreliability. Weisberg (2009) gives an example of this sort of evidence.

What you'd like to do in response to this sort of evidence is "rewind" to before you learned whatever it was the unreliable source told you, and then go through all your evidence since then and revise your attitudes in the light of it, without your belief state being tainted by the unreliable evidence. Whether such a procedure is practicable or even possible is something I shall not discuss further.

Testimony and higher-order evidence

This is related to the above kind of update. You interact with other agents and they will report their opinions to you. What are you to make of these reports? On some simple models, you will take the other agents to be perfectly truthful and this kind of learning will collapse into the first or second kind. But what if the reports are noisy? Or the agents possibly strategically untruthful? What sort of change in belief is appropriate in these circumstances? Under this heading we can group concerns about the *epistemology of disagreement* (Christensen 2009) as well as discussions of principles of *reflection* (van Fraassen 1984): think of your future self as another agent whose opinions you care about. We can also see work in Bayesian epistemology on unreliable signals as being an attempt to model this sort of update (Bovens and Hartmann 2003, Chapters 3 and 5).

Language change

Even more drastically, we can imagine modelling changes in formal language that describe the objects of belief. For instance, an agent might learn that "Superman" and "Clark Kent" are co-referring terms. Or that "The philosopher Nagel" could refer to either of at least three distinct persons. Or imagine someone at the beginning of the Nineteenth Century. Such a person would not have had any beliefs about the planet Neptune, since it wouldn't be discovered until the middle of the century. After its discovery in 1846, astronomers had to expand the set of propositions over which they had beliefs given all the new kinds of Neptune-related facts they now had.

4.6.2. Dilation

There are, as we have seen, a number of different ways you might try to update your credences on learning new evidence. In this section, I discuss some problems for generalised conditioning and related theories. I focus on generalised conditioning because it is the most common updating rule assumed in the literature and because it is closest in spirit to orthodox Bayesianism. It makes achieving precise Bayesianism in the precise limit easy. So these are problems for learning of the first type.

Imprecise probabilities suffer from two updating-related problems. The first of these – dilation – is well studied. This section will discuss it briefly. The next section deals with the less discussed problem of belief inertia. These are problems for generalised conditioning and related updating methods like the maximum likelihood method.

Dilation occurs when a precise representor becomes less precise on learning something. This seems worrying, since one would hope that evidence should decrease your ambiguity and thus sharpen your beliefs. Roger White's coin puzzle (White 2010) has been influential in the debate about imprecise credence in the philosophy literature, and so I feel I should outline and discuss his example. The example is, however, a little more complex than it needs to be, so once I have set out his example, I will introduce my own example which makes things clearer, I feel. The difference is a difference in presentation only.

White's coin puzzle

White's coin puzzle goes like this. I have a proposition P, about which you know nothing at all. I have written whichever is true out of P and $\neg P$ on the *Heads* side of a fair coin. I have painted over the coin so you can't see which side is which. I then flip the coin and it lands with the P uppermost. Using $\mathcal{P}$ to stand for your credences before the flip, and $\mathcal{P}_{\mathrm{P_{up}}}$ for your credences after having seen the coin land, and H_{up} and P_{up} for the propositions that the coin lands "Heads up" and "P up" respectively the following claims seem plausible.

1. $\mathcal{P}(P) = X$
2. $\mathcal{P}(H_{\mathrm{up}}) = \frac{1}{2}$
3. $\mathcal{P}_{\mathrm{P_{up}}}(P) = \mathcal{P}_{\mathrm{P_{up}}}(H_{\mathrm{up}})$
4. $\mathcal{P}_{\mathrm{P_{up}}}(P) = \mathcal{P}(P)$
5. $\mathcal{P}_{\mathrm{P_{up}}}(H_{\mathrm{up}}) = \mathcal{P}(H_{\mathrm{up}})$

The first is just a commitment to imprecise probabilism: if you know nothing about P, your belief in P should be maximally imprecise. Let X stand for whatever it is that fulfils this commitment. The second is just a commitment to have your beliefs guided by the objective chances. Third, seeing the coin land P uppermost, P_{up}, means that you now know that the coin landed Heads up if and only if P is true. That is, P_{up} is now equivalent to $H_{\mathrm{up}} \equiv P$. You should thus have the same beliefs in H_{up} and in P. The fourth and fifth items on this list just formalise the intuition that seeing the coin land P_{up} shouldn't change your opinions about the truth of P, nor about whether the coin landed Heads.

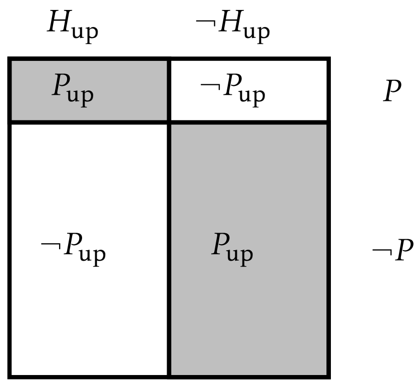

Figure 4.2.: White's coin example, visually (after Joyce 2011)

Figure 4.2 shows the situation for one particular belief about how likely P is. The horizontal line can shift up or down, depending on what the committee member we focus on believes about P.

From the above claims, we can show that X – the maximally imprecise belief state – can only be $\frac{1}{2}$ (Dodd forthcoming).

$$\begin{aligned} X &= \mathcal{P}(P) \\ &= \mathcal{P}_{\mathrm{P}_{\mathrm{up}}}(P) \\ &= \mathcal{P}_{\mathrm{P}_{\mathrm{up}}}(H_{\mathrm{up}}) \\ &= \mathcal{P}(H_{\mathrm{up}}) \\ &= \frac{1}{2} \end{aligned}$$

The figure makes this clear. The only way that learning that you are in the grey area would leave you with conditional probability of H_{up} of exactly a half is if P had prior probability $\frac{1}{2}$. It is only when the horizontal line is exactly in the middle that the proportion of grey on the P side of the horizontal line is a half. White's paper seems to argue in favour of some form of the principle of indifference, so perhaps this result isn't all that inimical to his views. However, I should think that many people will find this a strange outcome.

In order to diagnose this strangeness, let's consider a variation on White's puzzle. I have a stack of cards labelled "1" up to "100". On the back of the number 1 card I write whichever is true of P and $\neg P$. I write the false one on the back of all the other cards. I shuffle the cards and draw one at random. I show you that the drawn card has a P on it. Were White's above argument still valid in this example, the only possible value for $\mathcal{P}(P)$ would be $\frac{1}{100}$. However White's argument is not valid: premise 4 is false. It is not the case that $\mathcal{P}_{\mathrm{P}_{\mathrm{up}}}(P) = \mathcal{P}(P)$.

Seeing that a P-card is drawn at random is good evidence that P is false, since the odds favour a draw of a false proposition. This means that there is no requirement for the observation of P_{up} to be irrelevant in this situation. What this shows is that despite the irrelevance of the number to the truth of P, the way the game is set up can make a particular outcome good evidence for or against P. Note thus that White's premise 4 is only true because of an apparently unimportant feature of the set-up. A biased coin would undermine the argument. We can, however, build cases where dilation occurs even if we use a biased coin, but the initial belief in these cases cannot be perfectly precise. I will give an example in the next section, when I discuss the simpler set-up.

Now consider a case where there are only two cards, but you are initially pretty confident that P. Seeing the P-card drawn now looks like it should be evidence that the number 1 card has been drawn. This now goes against White's premise 5. Again, the game set-up makes it such that these apparently irrelevant instances of evidence *can* carry information. A little more precisely, consider the case where you *know* P to be true. Now seeing P_{up} is solid gold uncontrovertible evidence that H_{up}. Likewise, being very confident of P makes P_{up} very good evidence for H_{up}. If instead you were sure P was false, P_{up} would be solid gold evidence of H_{up}'s falsity. So it seems that $\mathcal{P}_{P_{\text{up}}}(H_{\text{up}})$ is proportional to prior belief in P. So let's go back to the original coin example and consider what it means to have an imprecise belief in P. Among other things, it means considering possible that P could be very likely. It is consistent with your belief state that P is such that if you knew what proposition P was, you would consider it very likely. In this case, P_{up} would be good evidence for H_{up}, and therefore premise 5 would be false. Note that in this case $\neg P_{\text{up}}$ would be just as good evidence against H_{up}. Likewise, P might be a proposition that you would have very low credence in, and thus P_{up} would be evidence *against* H_{up}. This shows that if you are taking the imprecise position seriously, you can't assume that P_{up} can't make a difference to your belief in H_{up}. Put differently, the point is that since $\mathcal{P}_{P_{\text{up}}}(H_{\text{up}})$ is proportional to prior belief in P, and your imprecise prior contains all the various priors for P, the posterior had better contain all the posterior beliefs in H_{up}.

It seems strange to argue that your belief in H_{up} should *dilate* from $\frac{1}{2}$ to $[0,1]$ upon learning P_{up}. It feels as if this should just be irrelevant to H_{up}. However, P_{up} is only really irrelevant to H_{up} when $\mathcal{P}(P) = \frac{1}{2}$. Any other precise belief you might have in P is such that P_{up} now affects your posterior belief in H_{up}. The imprecise probabilist takes into account *all the ways P_{up} might affect belief in H_{up}*.

Now I want to show that it is right that $\mathcal{P}_{P_{\text{up}}}(H_{\text{up}})$ should be equal to $\mathcal{P}(P)$. Consider some further principles:

6. For all $\mathbf{pr} \in \mathcal{P}$ we have $\mathbf{pr}(H_{\text{up}}|P) = \mathbf{pr}(H_{\text{up}})$ and therefore that $\mathbf{pr}(\neg H_{\text{up}}|\neg P) = \mathbf{pr}(\neg H_{\text{up}})$
7. $P_{\text{up}} \equiv (H_{\text{up}} \wedge P) \vee (\neg H_{\text{up}} \wedge \neg P)$
8. $H_{\text{up}} \wedge P_{\text{up}} \equiv H_{\text{up}} \wedge P$

Knowing just P doesn't change your beliefs in H_{up}. Their influence on each other is mediated through P_{up}, if you like. P and H_{up} are completely independent of each other until you know something about P_{up} which ties them together. Hence condition 6. 7 is just a description of the logical relationships between the events. It tells you how exactly P_{up} ties together the events P and H_{up}. This logic can be seen in Figure 4.2. 8 is a straightforward consequence of 7.

Now note that we can run the following argument for all $\mathbf{pr} \in \mathcal{P}$ since we know that $\mathbf{pr}(H_{\text{up}}) = \frac{1}{2}$.

$$\begin{aligned}
\mathbf{pr}(P_{\text{up}}) &= \mathbf{pr}(H_{\text{up}} \wedge P) + \mathbf{pr}(\neg H_{\text{up}} \wedge \neg P) \\
&= \mathbf{pr}(P)\,\mathbf{pr}(H_{\text{up}}|P) + \mathbf{pr}(\neg P)\,\mathbf{pr}(\neg H_{\text{up}}|\neg P) \\
&= \mathbf{pr}(P)\,\mathbf{pr}(H_{\text{up}}) + \mathbf{pr}(\neg P)\,\mathbf{pr}(\neg H_{\text{up}}) \\
&= \mathbf{pr}(P)\frac{1}{2} + (1 - \mathbf{pr}(P))\frac{1}{2} \\
&= \frac{1}{2}
\end{aligned}$$

Note that your belief in P seems to drop out of your belief in P_{up} which seems a little odd. But this is due to the game set-up, not due to any funny conditions I have imposed. The trick is that, as the above derivation makes clear, P_{up} is true when one of two conditions holds. These conditions are such that whenever one is unlikely, the other is likely and vice versa. With this fact in hand we can now do the following:

$$\begin{aligned}
\mathbf{pr}(H_{\text{up}}|P_{\text{up}}) &= \frac{\mathbf{pr}(H_{\text{up}} \wedge P_{\text{up}})}{\mathbf{pr}(P_{\text{up}})} \\
&= \frac{\mathbf{pr}(H_{\text{up}} \wedge P)}{\mathbf{pr}(P_{\text{up}})} && \text{by 8} \\
&= \frac{\mathbf{pr}(P)\,\mathbf{pr}(H_{\text{up}}|P)}{\frac{1}{2}} && \text{by the above argument} \\
&= 2\,\mathbf{pr}(P)\,\mathbf{pr}(H_{\text{up}}) && \text{by 6} \\
&= \mathbf{pr}(P)
\end{aligned}$$

Note that this result fits with the intuitive argument I gave earlier: if you have a strong belief in P, then learning P_{up} is going to make you more confident of H_{up}.

The above argument, coupled to White's condition 5 would show that $\mathbf{pr}(P) = \frac{1}{2}$ as before. I take this to show that White's condition 5 is unreasonable.

What I take this to show is that dilation is not *forced* on us by the reasonableness of claims 1 to 4, but that dilation is the *correct* response given the unknown value of the evidence P_{up} supplies about H_{up}. Given the reasonableness of conditions 6 and 7, dilation really is the *correct* response to the evidence P_{up}.

Even if we accept dilation as a fact of life for the imprecise probabilist, it is still *weird*. Even if all of the above argument is accepted, it still seems strange to say that your belief in H_{up} is dilated, *whatever you learn*. That is, whether you learn P_{up} or $\neg P_{\mathrm{up}}$, your posterior belief in H_{up} looks the same: $[0,1]$. Or rather, what it shows to be weird is that your initial credence was precise.

Marbles in urns

Now I want to introduce a slightly easier example of dilation. Imagine you know that there are a total of 10 black and 10 white marbles distributed somehow among the urns X and Y. Each urn contains 10 marbles. In other words, you know that $\mathbf{pr}(B|X) = 1 - \mathbf{pr}(B|Y)$ for all $\mathbf{pr} \in \mathcal{P}$. An urn will be selected at random by flipping a fair coin, and a marble drawn from it. Consider your credences before learning: $\mathcal{P}(W) = \{0.5\} = \mathcal{P}(B)$. You know that the number of white marbles and the number of black marbles are equal and that over the two urns their probabilities average out. It seems rational that your conditional credences are, however, imprecise: you know *nothing* about how the marbles are distributed between the urns, and so your belief representor, $\mathcal{P}$, plausibly includes probabilities that represent the possibility that urn X contains only white marbles and also functions that represent the possibility that X contains no white marbles, and everything in between. That is, $\mathcal{P}_X(W) = \mathcal{P}_X(B) = \mathcal{P}_Y(W) = \mathcal{P}_Y(B) = [0,1]$. So learning which urn is drawn from *dilates* your belief for white/black from $\{0.5\}$ to $[0,1]$. This illustrates the problem of dilation: updating on X or updating on Y both *dilate* your credence in B and in W from $\{0.5\}$ to $[0,1]$. That is, your beliefs in B and W get *less precise*, once you learn either X or Y. Note that this analysis of the problem makes an assumption about what the right belief after learning ought to be. It assumes the generalised conditioning strategy mentioned earlier.

Dilation seems like a problem. Why should your belief in B change on learning something that doesn't really affect what you know of the situation? Given that you don't know how X correlates with B, why should learning X make a difference to your belief in B? Of course the point is that your representor encodes all possible kinds of correlations between the events. It still seems strange that the prior belief is precise and the posterior belief is not, but note that *some* prior beliefs are imprecise in this case, for example $\mathcal{P}(B \wedge X)$.

I think, ultimately, dilation is a fact about imprecise probabilities that we must learn to live with. It is ubiquitous (Herron, Seidenfeld, and Wasserman 1994; Seidenfeld and Wasserman 1993). We must still do what we can to make dilation seem less strange. The first thing to point out is that dilation only seems odd because of the focus on $\mathcal{P}(B)$. Recall that $\mathcal{P}(B)$ is only a summary statistic: it is only a first approximation to what represents your belief in B.

Second, note that the prior belief is precise *only* because of the symmetry built in to the example. We know that, if X and Y are disjoint and exhaustive events, then for any event B,

$$\mathbf{pr}(B) = \mathbf{pr}(X)\,\mathbf{pr}(B|X) + \mathbf{pr}(Y)\,\mathbf{pr}(B|Y). \tag{4.5}$$

Moreover, for our particular marbles in urns example, the following symmetry also holds for all $\mathbf{pr} \in \mathcal{P}$:

$$\mathbf{pr}(B|X) = 1 - \mathbf{pr}(B|Y).$$

Putting these together, we see that, for all $\mathbf{pr} \in \mathcal{P}$,

$$\mathbf{pr}(B) = \mathbf{pr}(X)\,\mathbf{pr}(B|X) + \mathbf{pr}(Y)(1 - \mathbf{pr}(B|X)).$$

It now follows from the fact that $\mathbf{pr}(X) = 1/2 = \mathbf{pr}(Y)$ for all $\mathbf{pr} \in \mathcal{P}$ that $\mathbf{pr}(B) = 1/2$ for all $\mathbf{pr} \in \mathcal{P}$. This is how the symmetry of B given X, and B given Y, leads to a precise prior in our example. So each committee member "knows" how it will shift on learning X or Y. It just so happens that for each member who thinks B and X are positively correlated, there is one who thinks the correlation is negative to exactly the same degree. This ensures that the final belief, whether X or Y is learned looks "the same" at the level of $\mathcal{P}(B)$. Note that the weirdness is due to two things: the symmetry $\mathbf{pr}(B|Y) = 1 - \mathbf{pr}(B|X)$; and the fact that $\mathbf{pr}(X) = \frac{1}{2}$. If we didn't have the symmetry, it might be that, for example most of the marbles are in urn X and only one black marble is in urn Y. The probabilities would then be $\mathbf{pr}(B|X) = \frac{9}{19}$ and $\mathbf{pr}(B|Y) = 1$. Which would mean that $\mathcal{P}(B)$ would include $\mathbf{pr}(B) = \frac{28}{38}$. So the initial belief would not be precise. Since we could have swapped black and white around, we also have $\mathcal{P}(B)$ would include $\mathbf{pr}(B) = \frac{10}{38}$. As you increase the number of marbles, the range $\mathcal{P}(B)$ covers tends to $[\frac{1}{4}, \frac{3}{4}]$. Likewise, the precise initial belief also relies on the coin's being fair. Let's repeat the above calculation but without assuming $\mathbf{pr}(X) = \frac{1}{2}$ (but let's assume the symmetry holds).

$$\begin{aligned}\mathbf{pr}(B) &= \mathbf{pr}(X)\,\mathbf{pr}(B|X) + \mathbf{pr}(Y)\,\mathbf{pr}(B|Y)\\ &= \mathbf{pr}(X)\,\mathbf{pr}(B|X) + (1 - \mathbf{pr}(X))(1 - \mathbf{pr}(B|X))\\ &= \mathbf{pr}(B|X)[2\,\mathbf{pr}(X) - 1] + 1 - \mathbf{pr}(X)\end{aligned}$$

It is obvious from this that the $\mathbf{pr}(B|X)$ only drops out when $\mathbf{pr}(X) = \frac{1}{2}$. If $\mathbf{pr}(X) = \frac{1}{3}$ then

$$\mathbf{pr}(B) = \frac{\mathbf{pr}(B|X) + 1}{3}$$

So if, as in the original example, the proportions in the urns are completely unknown we have $\mathcal{P}(B) = [\frac{1}{3}, \frac{2}{3}]$. So dilation still occurs (since the conditional probabilities are still maximally imprecise) but its extent is less severe.

The standard examples of dilation in the literature (e.g. Seidenfeld and Wasserman 1993; Walley 1991; Wheeler ms.) typically take the form of known probabilities in two events having an unknown correlation; like the case discussed in section 4.4.2. The probabilities of X and Y are known, as are the (unconditional) probabilities of B and W. But the correlation between B and X is unknown, and this causes problems. In White's coin example, this structure isn't so clear. White's puzzle also perhaps trades on the intuition that fair coins land heads up about half the time, no matter what is learned. The literature on the Sleeping Beauty puzzle offers illustrations of how strong this intuition is (Bradley and Leitgeb 2006; Elga 2000; Schwarz ms.). For these reasons, I think that it is better to think about the current marbles-in-urns example.

Dilation doesn't seem quite so mysterious now. This example makes it clearer that $\mathcal{P}$ consists of a bunch of functions that encode specific opinions about how learning X would affect belief in B. White's example didn't make this clear and thus made dilation seem more mysterious than it should. It is a non-obvious aspect of the set up of White's coin puzzle that makes it the case that the updated credence in H_{up} *should* change unless the prior is exactly $\frac{1}{2}$. In the marbles in urns example, this link is made more obvious.

Reflection and dilation

It has been suggested that dilation examples show that imprecise probabilities violate the *Reflection Principle* (van Fraassen 1984). The argument goes as follows: "given that you know now that whether you learn X or you learn Y your credence in B will be $[0,1]$ (and you will certainly learn one or the other), your current credence should also be $[0,1]$." The general idea is that you should set your credences to what you expect your credences to be in the future. More specifically, your credence in B should be the expectation of your future possible credences in B over the things you might learn. Given that, for all the things you might learn in this example your credence in B would be the same, you should have that as your prior credence also. So having a precise credence in B to start with is irrational. Your prior $\mathcal{P}$ is not fully precise though. Consider $\mathcal{P}(B \wedge X)$. That is, the prior belief in the conjunction is imprecise. So the alleged problem with dilation and reflection is not as simple as "your precise belief becomes imprecise". The problem is "your precise belief *in B* becomes imprecise"; or rather, your precise belief in B *as represented by* $\mathcal{P}(B)$ becomes imprecise.

I think the issue with reflection is more basic. What exactly does reflection require of imprecise probabilists in this case?[41] Now, it is obviously the case that each credal committee member's prior credence is its expectation over the possible future evidence. That is exactly what equation (4.5) expresses. But somehow, it is felt, the *credal state as a whole* isn't sensitive to reflection in the way the principle requires. Each $\mathbf{pr} \in \mathcal{P}$ satisfies the principle, but the awkward symmetries of the problem conspire to make $\mathcal{P}$ as a whole violate the principle. This looks to be the case if we focus on $\mathcal{P}(B)$ as an adequate representation of that part of the belief state. But as I have said before, this is not an adequate way of understanding the credal state. $\mathcal{P}(B|X) \neq \mathcal{P}(B)$, but is this what reflection should require? Given that I think dilation is actually the reasonable response to the examples discussed, I think this suggests that we should look for a version of reflection that allows dilation.[42]

If you were keen on reflection and wanted *something* to respect the principle, then perhaps you would make the following definitions.

$$Q_{\mathbf{pr}}(X) = \left\{ \sum_i \mathbf{q}(X|E_i)\,\mathbf{pr}(E_i), \mathbf{q} \in \mathcal{P} \right\} \tag{4.6}$$

$$Q(X) = \left\{ Q_{\mathbf{pr}}(X), \mathbf{pr} \in \mathcal{P} \right\} \tag{4.7}$$

$Q_{\mathbf{pr}}(X)$ is what $\mathbf{pr}$ expects your credal state to be relative to learning some E_i from a partition. $Q_{\mathbf{pr}}$ is what committee member $\mathbf{pr}$ expects the committee belief to be after learning some E_i. That is, $\mathbf{pr}$ takes each committee member ($\mathbf{q}$) and works out what its beliefs would be on learning E_i. $\mathbf{pr}$ then takes the expectation over these (that is, expectation using her own beliefs $\mathbf{pr}$). $Q(X)$ is the set of the opinions

[41] See also Bradley and Steele (ms.[c]); Topey (2012).

[42] One man's modus tollens is another man's modus ponens...

of the **pr**s. You could then take this Q to be the right representation of belief. It will respect reflection and not be subject to dilation: $Q(B) = [0, 1]$. Even if worries about Q being sensitive to choice of partition $\{E_i\}$ could be overcome, I don't think this would be a particularly sensible response to dilation.

So dilation is a funny feature of imprecise updating. It is something imprecise probabilists must learn to deal with, but it is not a fatal flaw. Work still needs to be done to find out what sort of reasonable reflection principle holds of imprecise probabilities; what sort of representation of imprecise belief – that replaces $\mathcal{P}(B)$ – can be said to be properly reflective. Dilation might lead to some odd results in sequential decisions with learning, but that is beyond the scope of the current project.[43]

Dilation and the weight/balance distinction

It seems that examples of dilation undermine my earlier claim that imprecise probabilities allow you to represent the difference between the weight and balance of evidence: learning X appears to give rise to a belief which one would consider as representing *less evidence* since it is more spread out. This is so because the prior credence in the dilation case is precise, not through weight of evidence, but through this symmetry I discussed earlier. We cannot take narrowness of the interval $[\underline{\mathcal{P}}(X), \overline{\mathcal{P}}(X)]$ as a characterisation of weight of evidence since the interval can be narrow for reasons other than because lots of evidence has been accumulated. So my earlier remarks on weight/balance should not be read as the claim that imprecise probabilities can always represent the weight/balance distinction. What is true is that *there are* cases where imprecise probabilities can represent the distinction in a way that impacts on decision making. Dilation shows that I probably can't say much more than that.

4.6.3. Belief inertia

To introduce this second problem for imprecise probabilism, I want to return to an example similar to the one I discussed in relation to understanding the representor as a set of possible chances. This was based on Example 3.4.1 from Halpern (2003). Consider tossing a coin of unknown bias. Let's say that the proposition H_i encodes the history of observed tosses of the coin up to (but not including) the i^{th} toss. And let's say that X_i is the proposition "The i^{th} toss will land heads". So $\mathcal{P}(X_i|H_i)$

[43] But see Bradley and Steele (ms.[a]).

is the belief in the next coin landing heads given the history of the coin up to that point. Let's say that there are a number of chance hypotheses C_k on the table which each give some particular objective chance of the coin's landing heads. For each $\mathbf{pr} \in \mathcal{P}$ we have:

$$\mathbf{pr}(X_i) = \sum_k \mathbf{pr}(X_i|C_k)\,\mathbf{pr}(C_k) \tag{4.8}$$

$$\mathbf{pr}(X_i|H_i) = \sum_k \mathbf{pr}(X_i|C_k \wedge H_i)\,\mathbf{pr}(C_k|H_i) \tag{4.9}$$

Conditioning on the chance hypothesis screens off any correlation between H_i and X_i, because coin tosses are i.i.d. so:

$$\mathbf{pr}(X_i|C_k \wedge H_i) = \mathbf{pr}(X_i|C_k) \tag{4.10}$$

So any learning that happens goes through the effect of the history of tosses on the chance hypotheses $\mathbf{pr}(C_k|H_i)$. Fix some $\mathbf{pr}$, then by the Central Limit Theorem, the posterior probability converges to the chance almost certainly. This is true however biased against the true C_k the prior is (assuming it doesn't assign zero to the truth). That is, as i increases – as more coin tosses are observed – exactly one C_k has its posterior probability tend to 1. This is a good result. The precise Bayesian can learn.

And this is, of course, true for each probability in $\mathcal{P}$. However, there is a worry that despite this, there is a sense in which the imprecise probabilist might not be able to learn in the same way. Let's break down the important term in equation 4.9.

$$\mathbf{pr}(C_k|H_i) = \frac{\mathbf{pr}(H_i|C_k)\,\mathbf{pr}(C_k)}{\mathbf{pr}(H_i)} \tag{4.11}$$

Now, the $\mathbf{pr}(H_i|C_k)$ are determined by whatever that C_k says the objective chance of heads is. That's some basic combinatorics and the binomial distribution. So then things come down to this ratio of prior probabilities. If you know nothing about the chance hypotheses or the histories of coin tosses before any coins are tossed, then it seems reasonable that your representor contain functions that assign any allowable values to these sentences. In particular, $\mathcal{P}$ should contain functions that assign values to $\mathbf{pr}(C_k)$ and therefore, through equation 4.11 to $\mathbf{pr}(C_k|H_i)$ arbitrarily close to 0 and arbitrarily close to 1. Given the role $\mathbf{pr}(C_k|H_i)$ plays in learning about X_i, this entails that $\mathcal{P}(X_i|H_i) = (0,1)$, however big i is. That is, however much information you have gathered on the bias of the coin, there is some prior probability in your representor that is so confident that the chance is thus and so, that it won't be moved much by all the evidence you have amassed.

It might seem like I have said two contradictory things, but careful attention to quantifier order shows this is not true.[44] First I said "pick a **pr**, this **pr** will eventually converge to the chance". Second I said "pick some initial sequence of coin tosses H_i, at this time, there will be **pr** functions that have failed to converge, and indeed that have remained arbitrarily far from convergence".

The case can be seen clearly if we say there are only two chance hypotheses: $C_1 = \langle \mathbf{ch}(H) = \frac{1}{3} \rangle$ and $C_2 = \langle \mathbf{ch}(H) = \frac{2}{3} \rangle$. And let's say we have seen one hundred tosses, with approximately two thirds heads. This is, arguably good reason to believe C_2 and disbelieve C_1. However, if you started off being *fully non-committal* as to whether C_1 or C_2, then there are probabilities in your representor *arbitrarily close* to probability one in C_1. More specifically, fix H_i, now for any $\varepsilon > 0$, there will be some $\mathbf{pr} \in \mathcal{P}$ such that $\mathbf{pr}(C_2|H_i) < \varepsilon$. To see this all you need to do is make $\mathbf{pr}(C_1)$ close enough to 1. Now, it is also true that for this **pr**, with *enough evidence*, it will eventually begin to believe C_2. But for this new higher level of evidence there's an *even more* intransigent prior that was even closer to $\mathbf{pr}(C_1) = 1$ to begin with, and so on. The problem is that you're always stretching out the really crazy priors from the ends of the interval.

Thus, sadly, it seems that $\mathcal{P}(X_i|H_i) = \mathcal{P}(X_i)$ however much evidence is in H_i. So it seems that the imprecise probabilist cannot learn. Now, if we look at the credal committee members, they seem to be "bunching up" around one particular value, but there's always enough arbitrarily opinionated priors to always cover the same interval.

I don't claim to have a good solution to this problem, but let me note two things. First, there is something to this intuition that the committee members are "bunching up". Whatever measure you put over the set of probability functions – whatever "second order probability" you use – the "mass" of this measure gets more and more concentrated around the true chance hypothesis. Second, if your committee contained only finitely many members, then these problems would disappear. There would be a "most opinionated" member who would eventually believe the true chance. The problem arises in the infinite case because there are always more extreme, more opinionated priors. The endpoints don't move (if $\mathbf{pr}(C_1) = 1$, no amount of evidence can shift it) so no matter how much the posteriors get squished together around the true chance, there's always more

[44] A similar subtlety of quantifier order and what to hold fixed is well known in the phenomena of sensitive dependence on initial conditions and continuous dependence on initial conditions in chaos theory.

intransigent priors being stretched out from near the endpoints. Taking large finite samples of probabilities from the representor doesn't help with the problem of dilation, which can occur for finite sets of probabilities, so this isn't a full solution to the problems of updating.[45] If the prior representor's endpoints were not 0 and 1 but values strictly larger and smaller respectively, then the problem doesn't arise. This would be a very strong version of a principle of regularity.[46]

Maybe there is the beginnings of an answer to this problem of *belief inertia* in the above comments. Perhaps learning should be understood as the *movement* of the functions, rather than the spread of values of the posteriors. This doesn't offer much solace to the decision theorist, who would very much like to take certain bets on the coin toss and not others after learning some H_i, bets that she wouldn't have taken before learning. But if the sets of values for X_i are the same in each case, then how is this to be rationalised?

Sets of probability values, or rather, sets of expected values are going to play a big role in decision making. So it seems that just repeating the mantra "$\mathcal{P}(X)$ does not represent belief in X" can't be all there is to responding to these problems. Can we say something more helpful to the decision theorist? Here are a couple of observations. First note that $|\mathbf{pr}(X_i|H_i) - \frac{2}{3}| \leqslant |\mathbf{pr}(X_i) - \frac{2}{3}|$ for all $\mathbf{pr} \in \mathcal{P}$. So there is a sense in which the evidence does make certain bets look more favourable. Everyone agrees that a betting quotient of $\frac{2}{3}$ looks better after learning than it did before.

A second response to the problem of belief inertia is to consider what the "problematic" functions have in common: they are all outrageously opinionated at the start. Is it reasonable to be *so* confident that the chance is one third that seeing two thirds heads in a hundred tosses only leads you to assign probability ε to the proposition that the chance is two thirds? Maybe some kind of p-value-like "reasonableness cutoff" would allow you to ignore the crazy intransigent committee members? This response seems to somehow go against the inclusive, not-going-beyond-the-evidence roots of the imprecise probability framework. But if we take this to be "subjective" only in the sense that the choice of standard of evidence is subjective is, then this is perhaps enough to justify this response. The scoring rules approach to the reliability measure I discussed earlier in section 4.5.2 could help with this problem. Those extreme probabilities over the chance hypotheses

[45]The large finite sample approach will appear again in the next chapter, as one possible solution to the problems with decision making.

[46]The standard principle of regularity has it that only tautologies should get degree of belief 1 and only contradictions get 0.

are going to give rise to predictions that will have very low scores. If we use this as a criterion for reliability and then employ some sort of reliability cut-off, then they will drop out and the imprecise probabilist can learn after all. This happens, however low our reliability threshold is, so long as it does actually rule out some functions eventually. Perhaps there is something to the Maximum Likelihood method, or at least to the intuition behind it: you can safely ignore **pr** if **pr** makes the observed history very unlikely.

Wilson (2001) gives an example of another quirk where the posterior representor becomes more precise despite the evidence intuitively strongly supporting one hypothesis over the other. His solution is something like this second solution: excise the implausible priors.

A third response – and this is a response to both dilation and belief inertia – is to argue that what these examples show is that conditionalisation is not the right update rule. Perhaps the above problems speak in favour of going for some sort of recalibration update rule. Returning to the marbles in urns example of dilation, learning X does not give you any smaller reference class for which you have reliable statistics. That is, the "marble drawn from X" reference class (which we now know is the right one) is not a class for which we have good statistics. This means that the credence for B is still the proportion of black marbles in the "urn X or urn Y" reference class. Thus, $\mathcal{P}_X(B) = \frac{1}{2}$ for this kind of updating. The uncertainty regarding whether GC satisfies Reflection – or whether it even needs to satisfy Reflection – makes the worries regarding diachronic consistency of recalibration update rules less of a sticking point: all the update methods have problems on this front.

In any case, there is work still to be done on this topic. For now I want to move on to discussing how to make decisions using imprecise probabilities.

5. Imprecise decisions

> An economist is an expert who will know tomorrow why the things he predicted yesterday didn't happen today.
>
> *(Lawrence J. Peter)*

This chapter is the second half of the project I started in the last chapter: namely, to develop a conservative extension of the orthodox view on rational belief and decision under uncertainty. Such an extension is needed because of the difficulty that the standard view has with belief under severe uncertainty. This chapter focuses on decision making under severe uncertainty. A later chapter (chapter 7) discusses similar issues from a more practical angle.

We have now seen what imprecise probabilities are. This chapter asks the question "How should we make decisions with imprecise probabilities?" As we shall see, it is not nearly as simple as in the case of standard Bayesian decision theory.

After introducing the concept of a choice function, I situate imprecise choice between two other better-studied kinds of choice problem. I then discuss various kinds of imprecise choice and discuss their flaws. I use this discussion to motivate some criteria for rational imprecise choice.

5.1. The choice function

We have already met choice functions briefly in section 3.2.5 when we discussed Luce and Raiffa's representation theorem. We are going to be a little more careful here, as these functions will be central to this chapter. We think of a choice function as something that takes the set of available acts and outputs the set of choiceworthy acts. That is, it maps a set of acts into a subset of that set.

Definition 5.1.1 *A choice function is a function $\mathcal{C}\colon 2^{\mathbf{A}} \to 2^{\mathbf{A}}$ such that for all $A \subset \mathbf{A}$ we have $\mathcal{C}(A) \subseteq A$ and $\mathcal{C}(\mathcal{C}(A)) = \mathcal{C}(A)$.*

We interpret $\mathcal{C}(A)$ as being the set of choiceworthy acts among the feasible acts A. Call $\mathcal{C}(A)$ the "choice set". There are many ways of interpreting $\mathcal{C}(A)$.[1] A "Strong" interpretation would say that acts in $\mathcal{C}(A)$ are all equally the best act: there is nothing to choose between the acts in $\mathcal{C}(A)$ and you should be equally happy to take any of them. A weaker interpretation might be to say that all the acts in $\mathcal{C}(A)$ are better than the acts not in $\mathcal{C}(A)$. This interpretation does not preclude there being strict preference between the acts in $\mathcal{C}(A)$. An even weaker interpretation would be just to say that $\mathcal{C}(A)$ includes the best act, but to make no further assumptions about how acts in $\mathcal{C}(A)$ compare to those outside $\mathcal{C}(A)$ or among themselves.

I think that at least some of the scepticism about particular imprecise choice functions can be defused by being clear about what kind of interpretation we give the function. In precise Bayesianism, we can always take the strong interpretation of the choice function, but this interpretation might not always be applicable in the imprecise case.

Obviously, a weaker kind of choice function is less useful, but it may be that that is all we can achieve in this context. To return to a theme that runs all through this thesis, the kind of severe uncertainty at issue here might just preclude the determinate answers to decision theoretic questions we might wish to have. This relates directly to my weaker understanding of rationality. That is, perhaps in cases of severe uncertainty, rationality can only deliver us a weaker kind of choice set.

Rationality can provide necessary conditions for an act to be choiceworthy – or rather, sufficient conditions for an act to be unchoiceworthy – but it is harder to see why rationality should always provide necessary conditions for choiceworthiness. One criticism that was levelled at the imprecise probabilities approach in the last chapter was the claim that imprecise choice functions might fail to always determine an act that was optimal: they might fail to give you any advice. As I said when I discussed this earlier, such a criticism misses the point. To criticise imprecise choice for this flaw is to hold the theory to too high a standard. In cases of severe uncertainty, such determinate answers *shouldn't* always be available.

[1] A choice group talk by Conal Duddy on "shortlisting" inspired much of this discussion of interpretations of choice sets.

So perhaps it's best to keep two projects separate. First we want to know what rationality demands of imprecise decision; second we want to know how we should act in cases of imprecision. The answer to the former question might not fully determine an answer to the latter. But in those cases, we still need an answer to the second question. We will have to accept that the answer might not be fully rational. That's not to say that it will be *irrational* – violating rationality – but just that it will be *arational*: without rationality. Rationality can only get us so far. I leave the latter question until I discuss a specific example in the final chapter. I do this because I think the response to how to make decisions when rationality doesn't determine the answer will be context sensitive.

5.2. Kinds of decision

Here I situate the imprecise decision problem between two more familiar examples of decision problem. These examples differ only in the kind of belief representation you have. That is, we can think of the acts, outcomes and states being the same throughout. All that changes is the representation of belief over the states.

Example 3 Let's say I offer you the choice of two tickets c and d (on the understanding that you could choose neither n).

c	Gain £10 if X, lose £5 otherwise
d	Lose £5 if X, gain £10 otherwise
n	Gain £0 whatever happens

	X	$\neg X$
c	10	−5
d	−5	10
n	0	0

Which of these tickets should you pick? Well, you might say, it depends on what you think about X. The next three sections outline three different kinds of decision problems, involving three different kinds of attitude you might have to X. The three kinds of attitude are not fully distinct: one shades into the other, as we shall see.

5.2.1. Ignorance

Let's say you know *absolutely nothing* about X: you have no idea what sort of proposition it might be. In this instance you are in a situation of *decision under ignorance*. Decision theorists have been exploring this sort of problem for some time and the theory of decision under complete ignorance is pretty well covered (see Peterson 2009 chapter 3 or Luce and Raiffa 1989 chapter 13). So there are some possible events (in the example above, we have X and $\neg X$) and which obtains determines what your payout will be. These are the objects of belief, as we have seen before. You know nothing about which event is the case, so how ought you choose?

Probably the simplest decision rule is "Maximise minimum possible gain". I call this rule "Wald".[2] So you evaluate each act by its lowest possible outcome, and you act to maximise that. Let's define the worst and best outcomes that a can lead to as follows:

Definition 5.2.1 *For* $a \in \mathbf{A}$

- $\underline{a} = \min_X\{a(X)\}$
- $\overline{a} = \max_X\{a(X)\}$

Definition 5.2.2 $\text{Wald}(A) = \arg\max_{a \in A}\{\underline{a}\}$

This rule recommends going for n in Example 3. That is, $\text{Wald}(\{c, d, n\}) = \{n\}$. Is this a reasonable decision rule? Consider Example 4:

Example 4 You are offered the choice between these two acts:

a £100 if X, £1 otherwise

b £2 whatever happens

	X	$\neg X$
a	100	1
b	2	2

Wald sanctions taking b in this case. The following reasoning might support choosing a over b: "Unless X is very unlikely, a is the better bet." The point is that in genuine decision under ignorance, you can't rule out the possibility that X is very unlikely. So choosing a on this basis would be to make an unwarranted probabilistic assumption about how likely X could be. Criticism of particular ignorance choice rules often implicitly makes unwarranted assumptions about the probabilities of the various states. Some approaches we will see later basically amount to trying to make a principled probabilistic assumption in cases like this.

If Wald seems a rather pessimistic way of looking at things, perhaps the Hurwicz rule would suit better. It says "Maximise a weighted sum of maximum and minimum possible gains".

[2] I borrow this naming scheme from Milnor (1951)

Definition 5.2.3 $\text{Hurwicz}_\rho(A) = \arg\max_{a\in A}\{\rho\underline{a} + (1-\rho)\overline{a}\}$

If $\rho = 1$ then this is just the same as Wald. For all but the biggest values of ρ, Hurwicz opts for a in the decision problem in Example 4. In Example 3, for $\rho \leqslant \frac{2}{3}$, c and d are both chosen over n, otherwise n is chosen.

A different way to look at Example 4 is as follows: "even when b is better than a, the difference is small. But when a is the better act, it is *much* better. This points to another possible ignorance rule: Savage. Also known as "minimax regret", this rule says you should evaluate acts in terms of their "regret" and aim to minimise your maximum possible regret. Regret is defined as the difference between what you got from your act and what you *could have* got in that state by choosing a different act. So we first define **Reg** which gives us the maximum regret of an act a. Then Savage outputs the acts that minimise this quantity.

Definition 5.2.4 $\mathbf{Reg}(a) = \max_X \max_{b\in A}\{b(X) - a(X)\}$

Definition 5.2.5 $\text{Savage}(A) = \arg\min_{a\in A}\{\mathbf{Reg}(a)\}$

So for Example 4, the maximum regret of a is 1, because if X turns out false, you could have got 2 by choosing b rather than the 1 you actually got. b's maximum regret is 98 because if X is true, you could have got 100 from a rather than the 2 you actually did from b. Savage opts for n in Example 3.

Another way you might try and evaluate acts is not in terms of their regret, or in terms of the maximal and minimal outcomes, but in terms of the average outcome. This is the essence of Laplace. How this rule is normally formulated is to say "treat each state as equally likely, and maximise expectation". This comes out the same as saying "maximise average outcome". So assign probability of X and $\neg X$ to be one half each and then maximise expectation. Laplace effectively claims that there *is* a warranted probabilistic assumption in cases of ignorance. Laplace subsumes ignorance under the category of decision under uncertainty, which we shall turn to shortly.

There is some debate over which ignorance rule is the best. I find the debate somewhat strange: if you know *literally nothing* about the situation you find yourself in, why assume there should be *some* rationally determined course of action? Might it not be the case that you simply don't have enough information to make an informed choice? That's not to say that "anything goes". There are still unequivocally bad options even in decision under ignorance. For example, consider the case where we have acts a, b, c and act a is $b - 5$. That is, it is 5 worse than b for all states. It seems obvious that a should not be chosen. But this does not mean we are any closer to an answer as to which of b or c is better. I will return to this point later.

Before we move on, note that all of these rules satisfy the following property: if $a(X) \geqslant b(X)$ for all X, then b is not in the choice set. This property of not choosing acts that are *dominated* is typically something we would like to have for our choice rule.

5.2.2. Probabilistic expectation

Ignorance is one extreme of decision problems. To return to Example 3, the other extreme is where you know exactly the probability of X.[3] Perhaps you know that X is the proposition "The next flip of this biased coin will be heads". In this case, again, decision theory has you covered: *decision under risk* typically says you should maximise expected value.[4] Let's say you have examined this biased coin and found the probability of heads for it to be $\frac{4}{5}$. Under these circumstances c will

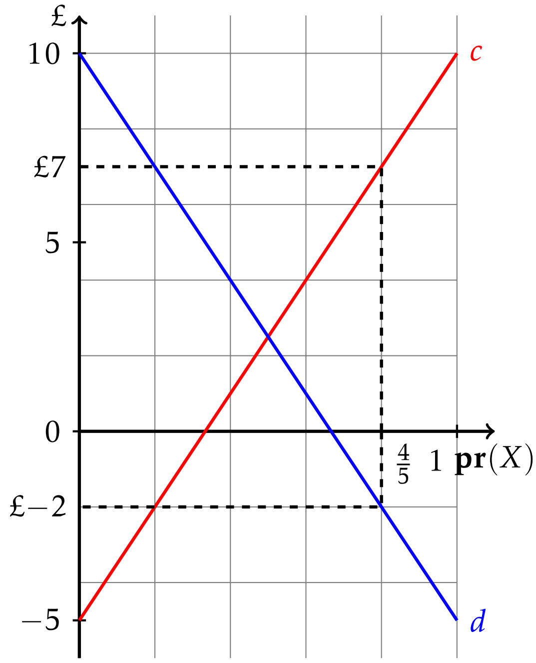

Figure 5.1.: c and d as functions of degree of belief

have expected value £7 and d will have £-2 (see Figure 5.1[5]). So c is a better bet

[3]Well, the other *extreme* is where you know whether or not X is true, but that is uninteresting for decision theory

[4]There are dissenting voices, I shall ignore them

[5]Graphs of expectation against probability appear in Seidenfeld (2004) and Steele (2007)

in this case. I don't have much to say about this case, except that it serves as a kind of "ideal limit" for imprecise decisions.

Recall that we defined probabilistic expectation in Definition 2.3.11 as:

$$\mathrm{E}_{\mathbf{pr}}(a) = \sum_i \mathbf{pr}(X_i)u(a(X_i))$$

So the above probabilistic expectation-based rule is then:

Definition 5.2.6 $\textsc{MaxExp}_{\mathbf{pr}}(A) = \arg\max_{a\in A}\left\{\mathrm{E}_{\mathbf{pr}}(a)\right\}$

The Laplace rule we saw in the previous section is a special case of this rule: the special case where the **pr** in question assigns equal probability to all states.

5.2.3. Imprecision

Now we come to imprecision. How are you to decide when you know *something* about X, but not enough to determine a probability distribution over the state space?

You could always resort to "decisions under ignorance" discussed above. But this seems unsatisfactory: if you know *something* about the probabilities, this information should feed into your decision making. This seems to be the guiding norm of decision theory: your decisions should be guided by your current state of knowledge. Ignoring the probability information you do have in your representor seems to violate this norm.[6]

You can't use the same maximising rule that precise probabilists can, because there's no guarantee that there will be one act that maximises expectation with respect to all probabilities in the representor. But when there is one act that maximises with respect to all probabilities, then you should certainly choose it. If all probabilities agree that a maximises, then a should be in the choice set. We could put it this way: a minimum requirement on imprecise choice is that if $\underline{\mathcal{E}}(a) \geqslant \overline{\mathcal{E}}(b)$ for all b, then a is optimal.[7]

To return to Example 3, which I discussed earlier, which of the bets c or d ought you take? Let's say your belief in X ranges over the interval $[0.2, 0.8]$. For each probability in your representor there is some expected value assigned to each event. For example, in Figure 5.1, the probability that assigned $\frac{4}{5}$ to X has expected value £7 for c and £-2 for d. We can consider the sets of expected values. For

[6]Indeed, on some interpretations of credence, it's impossible: that is, how you act is what determines what your beliefs are.

[7]Recall the definitions of these things from page 108.

each act, there is a range of expected values in the representor. Act c's expectation covers the range $[-2, 7]$ as does act d. Act n is such that all probabilities in the representor give it expected value £0 (see Figure 5.2).

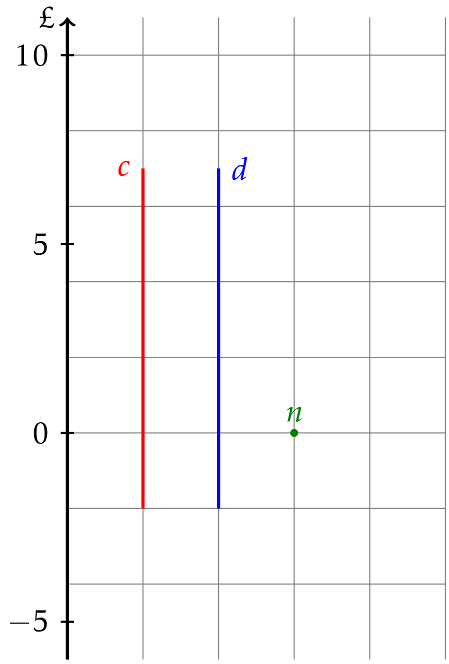

Figure 5.2.: The range of expected values

There is no clear best act in this scenario. So how best to make decisions? In the strict Bayesian model there is some act that *maximises* expected value. Or some collection of acts each of which maximises expectation. Figure 5.2 for the strict Bayesian would just be a series of dots, and it is obvious that the highest dot is the best. In the imprecise case, intervals can overlap, so what is "best" is not as clear. Acts whose expectations overlap are "incommensurable". It is not clear how to compare them to each other, thus evaluating them with respect to each other is difficult.

This imprecise belief model is, in a sense, "between" these two better studied modes of decision making. If there is just a single probability in your representor you'd better act like a maximiser. If your representor contains all possible probability functions – if all possible probability functions are compatible with your evidence – then you are effectively in a state of complete ignorance and you'd better act accordingly.[8]

[8] This is deliberately vague: whichever decision-under-ignorance rule you endorse, you'd better act *that* way when all probability measures are in your representor. Personally, I don't find any of the ignorance rules compelling, so maybe an approach to imprecise decision will inform the

Note that I have been restricting myself to choices that happen at a time. I will not be discussing *sequential choice* at all.[9] The static case will give us problems enough as it is.

5.3. How to think about imprecise choice functions

What does a reasonable imprecise choice rule look like? First, we need to be clear about what sort of rule we are interested in. Are we looking for one that admits of the strong interpretation discussed earlier? Or only some less strong interpretation? Second, we could wonder about necessary conditions on reasonable choice, or we could wonder about sufficient conditions.

There are many places in the literature where enterprises like this this have been developed. There are a great many ways we could approach the question of how best to settle on an imprecise decision rule. I survey some ways here.

One major source of inspiration is the literature on decision under ignorance. This will suggest some kinds of choice rule, as well as some of the conditions for reasonable imprecise choice. The most important sources here will be Milnor (1951) and Luce and Raiffa (1989).

Another important inspiration will be work on social choice theory: how to aggregate many individuals' preferences into a single group choice. If we think of each probability in your representor as a member of a credal committee that has to vote on what you should do, then the parallel between imprecise decision and social choice becomes clear. Here I will draw on Arrow's theorem (Gaertner 2009) and the work of Amartya Sen (Sen 1970, 1977).

There are two ways one might frame the discussion: in terms of an ordering over the acts (Arrow, Milnor), or in terms of a choice rule (Luce and Raiffa, Sen). I will talk in terms of choice rules, but we will see that relations will also play an important role.

There are two ways we could describe conditions on the choice function. One is just to put conditions on the functional form of the choice function. That is, we could impose intuitive conditions on the function with respect to how it interacts with unions and intersections of sets of acts. For example, is it reasonable to demand that if $B \subseteq A \subseteq \mathbf{A}$ then $\mathcal{C}(B) \subseteq \mathcal{C}(A)$? It turns out the answer is no. But that's an example of the kind of condition we might think to impose on a choice

choice of ignorance rule.

[9]But see Bradley and Steele (ms.[a],[b]); Steele (2010).

function.

There is another way we might want to impose constraints on reasonable choice functions. This is by restricting various kinds of relation associated with the choice function.

For this, we need some definitions.

Definition 5.3.1 *A choice function $\mathcal{C}$* pairwise satisfies *a relation $\succeq$ when, for all $a, b \in \mathbf{A}$:*

- *If $a \succeq b$ then $a \in \mathcal{C}(\{a, b\})$*
- *If $a \succ b$ then $\{a\} = \mathcal{C}(\{a, b\})$*

Definition 5.3.2 *A choice function $\mathcal{C}$* satisfies *a relation $\succeq$ when, for all $a, b \in A \subseteq \mathbf{A}$:*

- *If $a \succ b$ then $b \notin \mathcal{C}(A)$*
- *If $a \sim b$ then $a \in \mathcal{C}(A) \Leftrightarrow b \in \mathcal{C}(A)$*

We could then constrain reasonable choice by demanding that the choice function (pairwise) satisfies some particular relation defined on the acts. If C satisfies $\succeq$ then it pairwise satisfies it, but the converse need not be true.

A relation can also determine a kind of choice function.

Definition 5.3.3 *The* maximal set *for a relation $\succeq$ is $\mathcal{M}_{\succeq}$:*

$$\mathcal{M}_{\succeq}(A) = \{a \in A : \neg\exists b \in A, b \succ a\}$$

This $\mathcal{M}_{\succeq}$ is the set of acts that aren't strictly dispreferred to something else in the set.

Theorem 5.3.1 *(i) $\mathcal{M}_{\succeq}$ is a choice function and (ii) $\mathcal{M}_{\succeq}$ pairwise satisfies $\succeq$. (iii) If $\succeq$ is acyclic on A where A is finite then $\mathcal{M}_{\succeq}(A)$ is non-empty. (iv) Furthermore, if $\succeq$ is transitive, then $\mathcal{M}_{\succeq}$ satisfies $\succeq$.*

Proof (i) $\mathcal{M}_{\succeq}(A) \subseteq A$ by definition. It is equally obvious that $\mathcal{M}_{\succeq}(\mathcal{M}_{\succeq}(A)) = \mathcal{M}_{\succeq}(A)$.

(ii) We need to show that if $a \succeq b$ then $a \in \mathcal{M}_{\succeq}(\{a, b\})$. The only way a could fail to be in $\mathcal{M}_{\succeq}(\{a, b\})$ is if $b \succ a$. But this is ruled out by definition of $\succ$. If $a \succ b$ then $a \succeq b$, so by the above, we have that $a \in \mathcal{M}_{\succeq}(\{a, b\})$, and by definition, $b \notin \mathcal{M}_{\succeq}(\{a, b\})$.

(iii) Let $\geq$ be acyclic on some finite A. By a result similar to Lemma 1*l of Sen (1970, p. 16), acyclicity is sufficient to ensure non-emptiness of $\mathcal{M}_{\geq}(A)$.

(iv) If $a > b$ then $b \notin \mathcal{M}_{\geq}(A)$ by definition. Finally, assume for contradiction that $a \sim b$ and $a \in \mathcal{M}_{\geq}(A)$ but $b \notin \mathcal{M}_{\geq}(A)$. This means there exists some $c > b$. But $b \geq a$ so by transitivity[10] $c > a$, contradicting $a \in \mathcal{M}_{\geq}(A)$. ∎

Call a choice rule $\mathcal{C}$ *more discriminating* than $\mathcal{C}'$ when $\mathcal{C}(A) \subseteq \mathcal{C}'(A)$ for all A. $\mathcal{M}_{\geq}$ is the least discriminating choice function that satisfies $\geq$.

Theorem 5.3.2 *If $\mathcal{C}$ satisfies $\geq$ then $\mathcal{C}(A) \subseteq \mathcal{M}_{\geq}(A)$ for all A.*

Proof Let $a \in \mathcal{C}(A)$. Assume for contradiction that there is some $b \in A$ such that $b > a$. If there were such a b, then a would not have been in $\mathcal{C}(A)$ by definition of "satisfies". Thus $\neg\exists b \in A, b > a$. This is exactly the condition required for inclusion in $\mathcal{M}_{\geq}$. ∎

What if, instead of talking about maximality, we talked about *optimality*?

Definition 5.3.4 *The* optimal set *for a relation $\geq$ is:* $O_{\geq}(A) = \{a \in A : \forall b \in A, a \geq b\}$

What we will find is that optimality – which is stronger than maximality – is too strong a property. That is, $O_{\geq}$ is often empty. Consider the set $\{a, b\}$ where no preference holds between the two options. For this set, there are no optimal acts – although both acts are maximal. If the relation is complete, reflexive and acyclic then $O_{\geq}$ is nonempty (Sen 1977, p. 55).

In summary, we want to analyse what sort of choice rule makes sense for imprecise decision. We are going to proceed by imposing certain intuitive constraints on choice and showing that certain decision rules violate these principles. The principles will come in two flavours: restrictions on the functional form of $\mathcal{C}$ and relations that $\mathcal{C}$ must satisfy.

5.3.1. The basic requirements on imprecise choice

Here we want to put in place the absolutely basic, non-negotiable properties of decision rules. Let's consider a first possible decision rule. What about $\mathcal{C}(A) = A$. This satisfies the definition of a choice function. Unfortunately, it is thoroughly unhelpful! This choice rule is maximally permissive, or minimally discriminating. In general, we would like our choice rules to be as discriminating as possible.

[10] Strictly speaking, we don't really need transitivity here: we only need that $b \sim a$ and $c > b$ imply $c > a$.

This is something of a guiding principle for this chapter: *ceteris paribus*, more discriminating choice rules are better.

It seems obvious that if $\underline{\mathcal{E}}(a) \geqslant \overline{\mathcal{E}}(b)$ for all b, then a should be chosen. If a *cannot do worse* than *the best any other act can do*, then a is clearly the best act.

Definition 5.3.5 $ID(A) = \{a \in A : \forall b \in A \ \underline{\mathcal{E}}(a) \geqslant \overline{\mathcal{E}}(b)\}$

This is the decision rule that Henry Kyburg suggests in Kyburg (1983). It is called "Principle III".[11] Unfortunately, ID is often empty. That is, there is often no unambiguously best act in this sense.

This sort of failure to choose any choiceworthy acts seems a flaw in a choice rule. This suggests the first absolutely non-negotiable principle on choice functions:

Decisiveness *If* $\mathcal{C}(A) = \varnothing$ *then* $A = \varnothing$.

We need our choice functions to actually help us make choices. If the choice function can just "give up" then it isn't really that useful. There may be cases where "giving up" is in fact the right thing to do; for example in cases where you have to choose among acts of the form "n days in heaven followed by eternity in hell" for all $n \in \mathbb{N}$. It seems no choice of n is permissible, since there will always be a much larger n. Even if we want to grant this, the ID rule looks like it is definitely giving up too easily.

But we can use this *interval dominance* idea to further restrict reasonable choice rules: when some act does interval dominate all others, then the dominating act should be in the choice set.

Definition 5.3.6 $a \succ_{ID} b$ *iff* $\underline{\mathcal{E}}(a) \geqslant \overline{\mathcal{E}}(b)$

Note this is defined directly as an irreflexive relation, since it doesn't lend itself to having a reflexive part. This gives us another non-negotiable condition.[12]

Interval Dominance $\mathcal{C}$ *satisfies* $\succ_{ID}$

Often $\succ_{ID}$ is empty, so this condition will put no restrictions on choice. However, when it is not empty, the restrictions it puts on choice are reasonable.

So the plan in the following sections is to suggest some possible imprecise decision rules, including the main contenders from the literature, and show what conditions they violate.

[11] In response to Teddy Seidenfeld's comments (pp. 259–61), Kyburg changes his mind (p. 271). We will discuss this in due course.

[12] Note that $a \succ_{ID} b$ and $b \succ_{ID} a$ implies a and b have the same precise expectation. So the second condition of Definition 5.3.2 is still reasonable in this odd case.

5.4. Ignorance analogues

Now we move on to consider some imprecise decision rules built as "imprecise analogues" of our rules for decisions under complete ignorance.

5.4.1. Maximising lower and upper expectation

So our first decision rule is probably the simplest possible rule: "maximise $\underline{\mathcal{E}}$". We might term this $\mathcal{E}$Wald after the way it mirrors our "maximise minimum gain" rule (Wald) for decisions under ignorance.

Definition 5.4.1 $\mathcal{E}\text{Wald}(A) = \arg\max_{a\in A}\{\underline{\mathcal{E}}(a)\}$

This is sometimes described as "gamma-maximin" (Seidenfeld 2004). Like Wald in the ignorance case, this rule maybe seems overly pessimistic. This is the rule that Gärdenfors and Sahlin (1982) advocate.[13]

Maybe some rule that takes account of both the top and bottom of the interval would do the trick? Take an "ambiguous analogue" of the Hurwicz criterion:

Definition 5.4.2 $\mathcal{E}\text{Hurwicz}_\rho(A) = \arg\max_{a\in A}\{\rho\,\underline{\mathcal{E}}(a) + (1-\rho)\,\overline{\mathcal{E}}(a)\}$

This is actually a whole class of different decision rules depending on choice of ρ. If $\rho = 1$ then we recover maximise lower prevision ($\mathcal{E}$Wald). If a precise ρ value seems arbitrary, perhaps consider looking for acts that do well for many different values of ρ. Bandyopadhyay (1994) suggests a rule that, effectively, amounts to preferring a to b just in case a is better according to all values of ρ. This is obviously an incomplete relation, but it is more discriminating than Interval Dominance. However, there are some problems. $\mathcal{E}$Wald and $\mathcal{E}$Hurwicz suffer from the same problems, essentially and Bandyopadhyay's rule has the same flaws. Any decision rule that is sensitive to only the set of expected values looks like it will lead to problematic results.

Domination

Consider the following situation.[14]

[13] It might be said that given the slightly sophisticated account of the $\mathcal{P}$ that they give, my criticisms of $\mathcal{E}$Wald don't apply to these authors. I will leave off further discussion of their particular theory until later.

[14] For the examples in this chapter, unless otherwise specified, I assume that your representor is maximally uncertain. That is, I will assume $\mathcal{P}(X) = [0,1]$.

Example 5 There is a coin of unknown bias. you are offered the choice between these two bets:

f Win £1 if the next toss lands heads

g Win £1 if the next ten tosses all land heads

The first of these is clearly better. However, the set of expected values for each is $[0, 1]$ so a choice rule that only pays attention to the set of expectations will not be able to discriminate between these bets. Call an act b "dominated" if there is

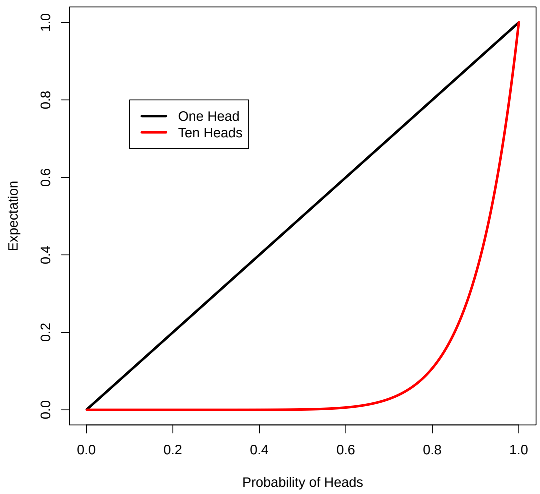

Figure 5.3.: Graphs of Example 5

another act a such that, for every member of the representor, a has at least as great an expectation as b.[15]

Definition 5.4.3 *b is* (weakly) dominated *by a if, for all* $\mathbf{pr} \in \mathcal{P}$ *we have* $\mathrm{E}_{\mathbf{pr}}(b) \leqslant \mathrm{E}_{\mathbf{pr}}(a)$*. Call b strongly dominated if the inequality is always strict.*

It seems reasonable that in this circumstance we should prefer a to b and in fact that we can safely ignore b when making our final decision.

For a fixed $\mathbf{pr} \in \mathcal{P}$, consider the relation $\succeq_{\mathrm{E}_{\mathbf{pr}}}$ defined by $a \succeq_{\mathrm{E}_{\mathbf{pr}}} b$ iff $\mathrm{E}_{\mathbf{pr}}(a) \geqslant \mathrm{E}_{\mathbf{pr}}(b)$. Each $\mathbf{pr}$ defines such a relation. If you like, $\succeq_{\mathrm{E}_{\mathbf{pr}}}$ is the relation that the function $\mathrm{E}_{\mathbf{pr}}$ represents. Consider $\succeq_{\mathrm{Dom}} = \bigcap_{\mathcal{P}} \succeq_{\mathrm{E}_{\mathbf{pr}}}$. Relations are often thought of as ordered pairs, so the set theoretic operation makes perfect sense. That is, the relation that holds between two acts only when every $\mathbf{pr} \in \mathcal{P}$ agrees about the

[15]It is normal to add the clause "and sometimes strictly greater" to this definition. This is to stop a being dominated by itself. Because of how I define $\succeq_{\mathrm{Dom}}$ in a moment, I don't do this.

direction of the inequality in their expectations. This is clearly exactly the relation of weak dominance. This motivates another important desideratum for imprecise choice.

Non-domination *$\mathcal{C}$ satisfies* $\succeq_{\text{Dom}}$

Note that this condition entails Interval Dominance. That is, when a interval dominates b, a dominates b. Put another way, $\succeq_{\text{Dom}} \supseteq \succeq_{ID}$.

The Sure Thing Principle

$\mathcal{E}$Wald and friends violate another property, known as the Sure Thing Principle (STP). Depending on what framework you work in, STP will be framed in a variety of ways.

The sure thing principle is often said to be violated by people's actual choices in games like those of the Ellsberg paradox. The property does, however, have some intuitive appeal, so it is worth discussing with a view to perhaps *limiting* if not eradicating violations of it. Luce and Raiffa's Axiom 8 corresponds to the "sure thing principle". Or, as Luce and Raiffa put it:

> Consider a probability mixture of two decision problems with the same actions and states. If the payoffs in the second problem do not depend on the act chosen, then the optimal set in the mixed problem is the same as in the first problem. Luce and Raiffa (1989, p. 290)

As we saw in section 3.2.3, Savage's version of STP (P2) talks only in terms of his "preference given X". This makes sense, since Savage doesn't, at this stage, have access to a probability function he can mix acts with. Broome (1991) characterises STP in terms of what he calls *Separability*. The idea is that how good sub-parts of the decision problem are can be evaluated in isolation. It guarantees that the evaluation of the acts will, in a certain sense, be an aggregation of the evaluations of the sub-parts. In the standard case that aggregation is done by probability-weighted sums.

Perhaps the best way to understand STP is with an example.

Example 6 I am going to ask you to choose a or b. Then I am going to flip a fair coin. If the coin lands heads, you gain £6 if $\neg X$, nothing otherwise. If it lands tails, a and b pay out as set out here:

a	Gain £10 if X, nothing otherwise			X	$\neg X$
b	Gain £2 if X, £8 otherwise		a	10	0
			b	2	8

The idea is that since what you choose – a or b – doesn't make a difference if the coin lands heads, then you should choose in order to get the better of the options when it matters (in the tails branch of the game).

$\mathcal{E}$Wald chooses b over a in the tails branch. But when you mix with the heads branch, a ends up looking better.[16] Given the intuitive force of the above reasoning behind STP, this violation does seem to speak against rules that violate it, like $\mathcal{E}$Wald.

Buchak (ms.) argues that violating STP is reasonable. Buchak's focus is on decision theories sensitive to *risk* – I made the simplifying assumption that you are risk neutral – but her argument is interesting nonetheless. She argues that the intuitive appeal of STP is actually intuitive appeal for only a weaker property: State-wise dominance. The claim is that STP is piggy-backing on the plausibility of this strictly weaker principle, and that in general STP is not plausible. State-wise dominance is the principle that "If $a(X) \geq b(X)$ for all X and that preference is strict at least once, then $a \succ b$". STP entails SWD but not vice versa. The argument is that the intuitive cases of STP are also cases of SWD, and that cases of STP that aren't also cases of SWD are not as intuitive. This undermines the claim that STP is intuitive. However, the above example is not a case of SWD but it is a case of STP. It seems that there is something intuitive about the principle even in this non-SWD case. So I don't see the force of Buchak's argument. I appreciate that this is just trading on intuitions and so isn't a very strong argument.

Everything I say to criticise these sorts of simple $\mathcal{E}$Wald type rules applies also to decision theory built up from some other kinds of representation of uncertainty. For example, there is some literature in economics on maximising Choquet-expected utility (Eichenberger and Kelsey 2009). In this sort of work – typically in the behaviourist/representationalist paradigm I have criticised elsewhere – the ultimate conclusion is that maximising Choquet expected utility is rational. Recall a Choquet capacity is analogous to a lower probability function in that it represents a lower bound on estimates of the probability. This framework is less expressive than sets of probabilities: it lacks the resources to express the

[16] That is, the payouts of a and b for the "mixed" decision problem are "5 if X, 3 otherwise" and "1 if X, 4 otherwise" respectively.

structural constraints on beliefs like "betting on one head is always at least as good as betting on ten heads". Ignoring this dominance reasoning is the flaw both with Choquet-expected utility and $\mathcal{E}$Wald in the current setting.[17]

5.4.2. Minimising regret

What about an $\mathcal{E}$Savage rule? For example, we could define an imprecise analogue of "regret" as follows:

Definition 5.4.4 $\mathcal{E}\mathbf{Reg}(b) = \max_{\mathbf{pr}\in\mathcal{P}}\left\{\max_a\{\mathrm{E}_{\mathbf{pr}}(a)\} - \mathrm{E}_{\mathbf{pr}}(b)\right\}$

For an act b, we look at how far b is from the act that performs best for that probability function. We do this for all probabilities, and take the maximum of these "probabilistic regrets". We then try to minimise this. That is, for each probability in your representor, subtract the expectation of act b from the "top act" for that probability.

Definition 5.4.5 $\mathcal{E}\mathrm{Savage}(a) = \arg\min_{a\in A}\{\mathcal{E}\mathbf{Reg}(a)\}$

The maximum of those values over all probabilities in your representor is your maximum regret for act b. you should act to minimise that.

This at least gets it right for Example 5. The single heads bet has regret 0 everywhere, while the ten heads bet has positive regret everywhere except $\mathbf{pr}(H) = 0$ and $\mathbf{pr}(H) = 1$. This rule is sensitive to important aspects of the structure of the belief representor that get lost when thinking about $\underline{\mathcal{E}}$ and $\overline{\mathcal{E}}$ only.

Unfortunately, $\mathcal{E}$Savage violates STP,[18] as well as some even more important properties. Consider the following scenario. You go to a restaurant and see that the menu consists of Fish, Steak or Chicken. You decide on Chicken. The waiter comes to take your order and tells you there is no more Fish. So you decide to have the Steak. This story seems a little odd. Why should the availability of an option you don't choose cause a switch in choice like the one exhibited in the move from Chicken to Steak? It seems like a reasonable choice rule should be somewhat consistent under various kinds of expansion or contraction of the option set.[19]

---

[17]See also the partition-sensitivity worry I mentioned in section 4.5.3.

[18]For example, in Example 6, both acts have the same regret in the tails branch, but a has lower regret in the mixed game. So a minor modification of this game would demonstrate a proper violation.

[19]Imagine you are at a sushi restaurant. You prefer salmon over tuna, but find that salmon needs to be prepared by a good chef in order to be tasty. So you take the safe option and go for tuna. Now you notice that the restaurant offers fugu – pufferfish – which is poisonous unless

There are many different properties of varying strengths in this domain. Arrow had "Independence of Irrelevant Alternatives" which, among other things, entails a form of expansion and contraction consistency. Sen has his alpha and beta (etc.) properties, which we will see in due course. Luce and Raiffa have their axioms 6 and 7 and its strengthenings. Milnor had "row adjunction" and "special row adjunction". All of these will be fitted into this framework.

Consider contraction of the option set first.

Contraction Consistency $\quad C(A \cup B) \subseteq \mathcal{C}(A) \cup C(B)$

This rule is more normally seen in one of these equivalent forms:

$$\text{If } a \in \mathcal{C}(A), B \subseteq A, a \in B \text{ then } a \in \mathcal{C}(B) \tag{5.1}$$

$$\text{If } a \notin \mathcal{C}(B), B \subseteq A, a \in B \text{ then } a \notin \mathcal{C}(A) \tag{5.2}$$

Think about moving from A to B. The idea is that if it was good enough in A, and it's still an option, it should still be good enough in B. This property is also known as Sen's alpha condition (Sen 1970, 1977). I am following Gaertner (2009) in calling it "contraction consistency", but it also somewhat restricts expansion of the option set also. Luce and Raiffa have a version of (5.2) as their Axiom 7.

$\mathcal{E}$Savage violates this property. Given that the assessment of each act depends on maxima over the set of acts, it shouldn't be surprising that adding or removing acts can affect how the acts are evaluated.

Example 7 Consider the following acts, and consider choosing when h is an option, and when it isn't.

f Gain 10 if X, nothing otherwise

g Gain 8 if X, 1 otherwise

h Gain 4 if X, 3 otherwise

	X	$\neg X$
f	10	0
g	8	1
h	4	3

Now, in a choice between f and g, f wins, since it is at most 1 from the best act (when $\mathbf{pr}(X) = 0$) while g drops to 2 less than the top act (when $\mathbf{pr}(X) = 1$). All this changes when h becomes an option. Now, when $\mathbf{pr}(X) = 0$, f is 3 from the top

prepared properly. You change your order to salmon. This seems like a failure of Contraction Consistency, but it isn't really. Learning that the restaurant offers fugu gives you evidence that they have a good chef. This changes *each* of the acts on offer, so the change isn't just a simple addition of a new act. Thus the axiom doesn't apply. Thanks to Nick Baigent for this example.

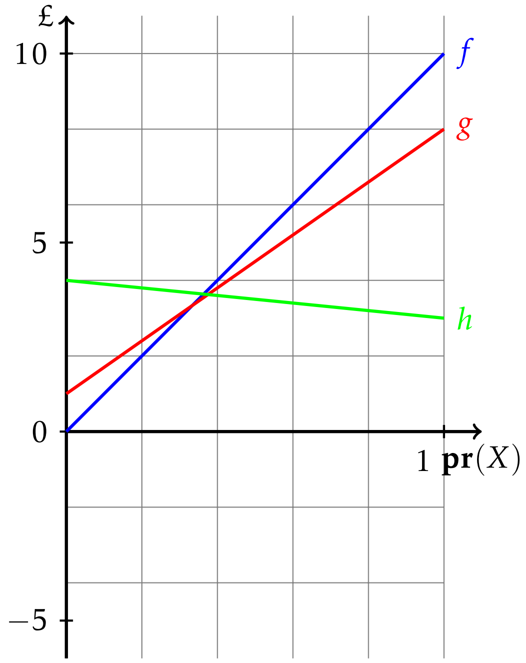

Figure 5.4.: Example 7 as a graph of expectation against probability

act while g's maximum distance from the top stays at 2. So the addition of h has switched the evaluation of the two acts round! In short, $\mathcal{E}$SAVAGE$(\{f, g, h\}) = \{g\}$ but $\mathcal{E}$SAVAGE$(\{f, g\}) = \{f\}$. This violates CONTRACTION CONSISTENCY.

We have seen that adding acts to the choice set should not be able to make a non-choiceworthy act into a choiceworthy one, at least not without giving you extra information about the choice problem. What about the other way? When should a choiceworthy act be made no longer choiceworthy by the addition of new acts?

Let's consider an example of an expansion where turning a choiceworthy act non-choiceworthy seems reasonable. A new act can make some old choiceworthy act no longer choiceworthy. That is, if I add b to a decision problem where a used to be the best, and b is better than a, then a is no longer choiceworthy. So perhaps we should demand that the new act's being choiceworthy is a necessary condition for the old act changing from optimal to non-optimal. Let's call this "strong expansion consistency". This is Luce and Raiffa's first strengthening of Axiom 7. They call it Axiom 7′.

> STEC *If $a \in \mathcal{C}(A)$ but $a \notin \mathcal{C}(A \cup B)$ then there is some $b \in B$ such that $b \in \mathcal{C}(A \cup B)$*

$\mathcal{E}$SAVAGE also violates this stronger property.

5.4.3. Equivocating

What about the fourth kind of decision under ignorance rule? Is there any imprecise analogue of Laplace? What we need to do in the ignorance case is construct a probability function that makes each state equally likely and maximise expectation with that. What we need to do in this case is pick out some particular probability function depending on your representor, and use this to maximise expectation.

I've already criticised the Objective Bayesian norm of Equivocation, which is effectively a method for picking out a probability to use.

Another approach with some of these characteristics is the "Transferable Belief Model". Smets and Kennes (1994) argue that you should maintain a set of probabilities in a "credal set" but that when called on to make a decision, you should collapse this into a "pignistic probability" and maximise expectation with respect to that.

In what follows I have less to say about this sort of approach. It seems there are epistemological problems with deciding how to "average" your probability judgements, but once that is achieved, the decision theory is wholly orthodox. I have argued that representing the level of your uncertainty is important, so to then have your decision theory "average away" all the severe uncertainty would be counterproductive. So I focus my attention on properly imprecise decision theories.

However, it might be that $\mathrm{E}_{\Downarrow\mathcal{P}}(f)$ – the expected value of f by the lights of the maximally equivocal probability – is a component in a more complex choice rule. I will discuss one such rule later.

5.5. Ruling out acts

All of the above rules basically work the same way: they determine some quantity that measures how good an act is, and then they recommend trying to pick acts that maximise that quantity. So they output strong choice sets, in the sense that all the acts in the choice set come out the same on the measure of choiceworthiness. They try to find some quantity that determines a sufficient condition for an act's being reasonable. But this seems a little strong. Given the kind of ambiguity we are dealing with, Who's to say that there is some metric that always delivers the optimal act?

Perhaps it is safer to take a different tack to constructing decision rules. Let's look at restricting the space of reasonable acts. Instead of considering which acts are the "best" in some sense, I want to focus on which acts you can justifiably rule out. I want to look at the necessary conditions on an act's being good. This gives us a choice set that only admits of a weaker interpretation. That is, all we can say of the acts in the choice set is that they are all better than those that have been excluded. But we cannot say that the acts remaining in the choice set are necessarily all choiceworthy (or comparable).

5.5.1. Non-domination

Recall that in Example 5 we criticised rules that focus only on the set of expectations for not taking into account that one act is always better than the other whatever the chance of heads. This suggested our Non-domination condition. So we want our choice function to satisfy this dominance relation. Let's explore the rule that satisfies this property, but is otherwise undiscriminating. That is, choose only among acts that have no act that is always better.

Definition 5.5.1 $\mathcal{U}(A) = \mathcal{M}_{\geq_{\mathrm{Dom}}}(A)$

This rule satisfies Non-domination, obviously, but also Contraction Consistency and STP.

Non-domination is, arguably, too weak a requirement however: many intuitively unreasonable acts are nevertheless undominated. $\mathcal{U}$ is not discriminating enough.

Non-domination also has a couple of minor problems.

All-or-nothing expansion consistency

First, let's look at an unproblematic violation of a property found in the literature. I shall call it "all-or-nothing expansion consistency". It is called $7''$ by Luce and Raiffa and "beta" by Sen. This says that if an old choiceworthy act is made non-choiceworthy, then *all* old choiceworthy acts are made non-choiceworthy. As Luce and Raiffa show, this "all-or-nothing" character of $7''$ makes sense only when you are evaluating the acts on a single scale.

All-or nothing *If $a \in \mathcal{C}(A)$ but $a \notin \mathcal{C}(A \cup B)$ then, for all $b \in \mathcal{C}(A)$, we have $b \notin \mathcal{C}(A \cup B)$*

Here is an example of how $\mathcal{U}$ fails all-or-nothing expansion consistency.

Example 8 Consider the choice between f and g, and the choice between f, g and h.

f Gain £10 if X, nothing otherwise

g Gain nothing if X, £10 otherwise

h Gain £11 if X, £1 otherwise

h dominates f, so in the expanded decision problem, f is not optimal. However, g is still undominated, so this violates all-or-nothing.

Imagine you are picking your team for the school gymnastic competition. You need the best long jumper and the best high jumper in class A. So you have a choice function that picks out Laura as the best long jumper and Horatio as the best high jumper. That is, the choice function "best athletes" chooses these two students. Now imagine that we expand the set of available students to be the whole year group not just the class. As it happens, Herbert, in class B is a better high jumper than Horatio. But Laura is the best long jumper in the year. So the choice rule now picks out Laura and Herbert for the team. This rule violates the above all-or-nothing rule, because different members of the choice set are included based on different kinds of evaluations. The choice function for each sport will satisfy this condition, but the overall "best athletes" function does not. Such a choice function can't be given a strong interpretation. I claim that imprecise decision can be a little like this, and thus that all-or-nothing should not be required. So it's no flaw in $\mathcal{U}$ that it violates this property.

Sugden (1985) discusses a similar example where one race car is faster and another is more manoeuvrable: the first will win in a head to head race, but the second will win if there are other cars on the track. Thus the "race winning function", if you like, does not satisfy this property.

Continuity

First consider a minor modification on an example we saw earlier (Example 5).

Example 9 There is a coin of unknown bias. You are offered the choice between these bets:

f Win £1 if the next toss lands heads

g Win £1 + ε if the next ten tosses all land heads win £ε otherwise

Now g is no longer dominated, but it is still, intuitively, a bad deal. Unless the probability of heads is pretty extreme, you're better off with f. One might phrase the intuition this way: g is only ε away from being dominated. The smaller ε gets, the less reasonable this act seems. One might consider it a flaw in a decision

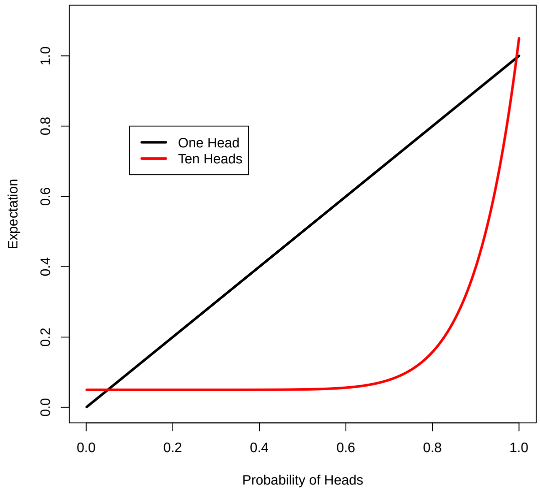

Figure 5.5.: A graph of Example 9

function that it allows an act to be in the choice set no matter how close it gets to some limit, but still make the "limit act" impermissible.

This motivates a property termed *continuity*. Continuity is about sequences of decision problems. A sequence of decision problems is defined by a collection of sequences of outcomes. I borrow Ken Binmore's gloss on Milnor's "continuity" axiom here:

> Consider a sequence of decision problems with the same acts and states, in all of which $a_j \succ a_i$. If the sequence of matrices of outcomes converges, then its limiting value defines a new decision problem in which $a_j \succeq a_i$. Binmore (2008, p. 137)

Continuity says that if we keep changing the payoffs slightly and the preference remains stable, then in the limit of our fiddling, the preference will still be stable. The change from "$\succ$" to "$\succeq$" is to allow this to be consistent with a case where act a_j tends to a_i. Or rather, each consequence of a_j tends to the corresponding consequence of a_i. In the limit, they are equal, so there cannot be strict preference between them. Or in terms of optimal acts, if we keep changing the payoffs slightly and the set of optimal acts remains stable, then in the limit, the set of optimal acts is stable.

Note that the slightly less discriminating choice rule "choose only among the non-strongly-dominated acts" does satisfy continuity. That is, a sequence of undominated acts can have a weakly dominated limit, but not a strongly dominated one.[20]

Convexity

I don't really have an intuition for the following property, but it appears in a number of places, so it is worth discussing briefly.

Convexity says that "if $a_1, a_2 \in \mathcal{C}(A)$ then $xa_1 + (1-x)a_2 \in \mathcal{C}(A)$ for $x \in [0,1]$". This is Luce and Raiffa's Axiom 9. Milnor has a version of this that says that mixtures of acts are weakly preferred. That is, it says "if $a_1 \sim a_2$ then $xa_1 + (1-x)a_2 \succeq a_1, a_2$ for $x \in [0,1]$." Of course, this only makes sense if these "mixtures" are in A.

A mixture of undominated acts can be dominated (See Table 5.1). Now each of a_1 and a_2 are undominated. But the mixture is dominated by a_3. So convexity is not true for $\mathcal{U}$.

	s_1	s_2
a_1	2	−2
a_2	−2	2
a_3	1	1
$0.5a_1 + 0.5a_2 = a_4$	0	0

Table 5.1.: A mixture of undominated acts can be dominated

Would it be possible to construct a rule that never picks an act whose mixtures with other acts can be dominated? Would it be desirable?

I find myself untroubled by violations of convexity. So why might we consider it a condition of rational decision? There is an idea that a mixture is at least as good as the worst part of the mixture, and at most as good as the best part. This is related to what Bradley (2007) calls the "Averaging Slogan". In particular, if both parts of the mixture are good enough to make it into $\mathcal{C}(A)$, then so should the mixture be. But this seems to be trading on the same "single-criterion" intuition as the above "All-or-nothing expansion consistency". Say Horatio is a terrible long-jumper and Laura a terrible high-jumper. Someone who was a "mixture" between these people – call him/her Lauratio – would not be as good as them. If Lauratio is an average high-jumper and an average long-jumper, then she/he doesn't deserve a place on the athletics team.

[20] The ten heads plus epsilon act in Example 9 is not ruled out by this less discriminating rule either.

5.5.2. E-admissibility

Another restriction of the act set – "E-admissibility" – is due to Isaac Levi (Levi 1974, 1986). An act is E-admissible if there is some probability in your representor such that that act maximises expectation with respect to that probability function. In picture terms an act is E-admissible if it is "on top" for some probability value in the representor. Levi argues that you should only choose among E-admissible acts.

Definition 5.5.2 $L(A) = \bigcup_{\mathbf{pr}} \mathcal{M}_{\geq_{\mathrm{E}_{\mathbf{pr}}}}(A)$

This might be more perspicuously rephrased as:

$$L(A) = \{a \in A : \exists\, \mathbf{pr} \in \mathcal{P}, \forall b \in A, \mathrm{E}_{\mathbf{pr}}(a) \geqslant \mathrm{E}_{\mathbf{pr}}(b)\} \tag{5.3}$$

As it stands, the definition of E-admissible isn't quite good enough. Recall Example 5 where we had the choice between a bet on heads and a bet on ten heads in a row. The latter maximises expectation for $\mathbf{pr}(H) = 0$ and $\mathbf{pr}(H) = 1$ so it is E-admissible. This act is, however, weakly dominated. I expect that Levi would want to rule out choosing weakly dominated acts, even if they maximise for some $\mathbf{pr}$. We could do two things here. We could either weaken our definition of $\mathcal{U}$ so that only strongly dominated acts are thrown out, or we could give Levi the benefit of the doubt and say that what he was really interested in was $L(\mathcal{U}(A))$. We shall call this $\mathcal{L}(A)$.

We know that $\mathcal{L}(A) \subseteq \mathcal{U}(A)$. There are undominated acts that are not E-admissible, for example n in Example 3. So we in fact know that $\mathcal{L}(A) \subset \mathcal{U}(A)$ for some A. So $\mathcal{L}$ is more discriminating that $\mathcal{U}$.

One motivation that might be offered for preferring $\mathcal{L}$ to $\mathcal{U}$ is that undominated acts can be dominated by mixtures of other acts, but E-admissible acts can't be dominated by mixed acts. For example, n in Example 3 is undominated but dominated by a 50:50 mixture of f and g. No E-admissible act can be so dominated. This seems like a strange argument. If the mixed acts are available acts, then n won't be undominated. If the mixed acts aren't in the set of available acts, then it's irrelevant that there is something that dominates it. A second point against this is that not all undominated acts can be dominated in this way. In a moment, we will see such an undominated act that can't be dominated by mixtures. The "Same" act in Example 10 is one such act.[21]

[21] More carefully, fix a set of acts A. Only some undominated acts in A are dominated by mixtures of acts in A. We could still have that the act is undominated in some larger set of acts A' such that it is dominated by mixtures of acts in A'.

Given that E-admissibility is more discriminating and given that non-domination is arguably too permissive (not discriminating enough), one might think that E-admissibility is obviously the better rule. But $\mathcal{L}$ seems to "get things wrong" in certain cases, where $\mathcal{U}$ gets it right. That is, $\mathcal{L}$ rules out some intuitively reasonable acts as we shall see shortly. $\mathcal{L}$ doesn't help solve any of the problems with $\mathcal{U}$. That is, $\mathcal{L}$ also violates continuity and convexity as well as sanctioning betting on ten heads in Example 5.

Union consistency

$\mathcal{L}$ violates the following reasonable seeming property.

Union consistency $\quad \mathcal{C}(A) \cap \mathcal{C}(B) \subseteq \mathcal{C}(A \cup B)$

This is Sen's gamma condition. It is sometimes seen in this equivalent form:

$$\text{If } a \in \mathcal{C}(A), a \in \mathcal{C}(B) \text{ then } a \in \mathcal{C}(A \cup B) \tag{5.4}$$

The motivation here is that if you would choose Steak out of Steak or Fish, and you'd choose Steak out of Steak or Chicken, then you should choose Steak when all three options are on the menu. $\mathcal{L}$ violates this property, as can be seen from considering Example 3. $\mathcal{L}(\{f,n\}) = \{f,n\}$ and $\mathcal{L}(\{g,n\}) = \{g,n\}$, but $\mathcal{L}(\{f,g,n\}) = \{f,g\}$.

As I mentioned earlier, Kyburg (1983) suggests his Principle III, which is essentially what we called Interval Dominance. In the commentary on Kyburg's paper, Teddy Seidenfeld suggested two improvements on Principle III which are effectively $\mathcal{U}$ and $\mathcal{L}$. Seidenfeld notes that $\mathcal{L}$ violates Union consistency (although he mistakenly refers to it as Sen's beta condition). Ultimately, Seidenfeld suggests that despite this failure $\mathcal{L}$ is the better improvement on Principle III (Kyburg 1983, pp. 259–61). In Kyburg's response (p. 271) he concedes that Interval Dominance should be improved on, but finds the violation of the above property to be undesirable, and thus opts for $\mathcal{U}$. Seidenfeld (2004) points out that $\mathcal{L}$ violates Union consistency and also All-or nothing.

Intuition failure

E-admissibility also just seems to get it wrong in certain intuitive cases.

Example 10 You are betting on a coin of unknown bias. It will be tossed twice. The following bets are available to you.

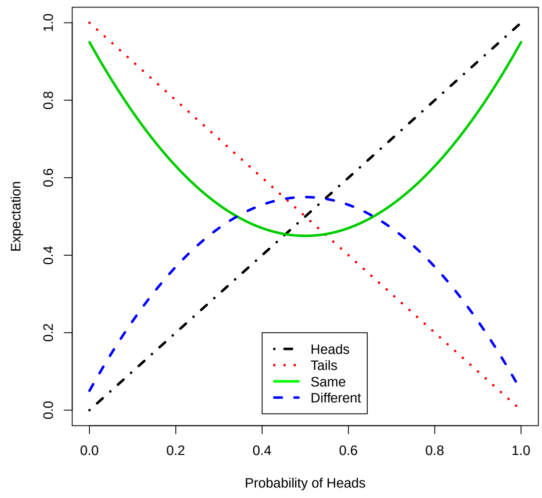

Figure 5.6.: The decision problem of Example 10

SAME Win £1 $-\varepsilon$ if the coin lands the same way twice, lose ε otherwise

DIFFERENT Win £1 $+\varepsilon$ if the coin lands on different sides each time, win ε otherwise

HEADS Win £1 if the first toss lands heads, 0 otherwise

TAILS Win £1 if the first toss lands tails, 0 otherwise

	HH	HT	TH	TT
S	$1-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$1-\varepsilon$
D	ε	$1+\varepsilon$	$1+\varepsilon$	ε
H	1	1	0	0
T	0	0	1	1

The intuition here is that whatever value **pr** takes, S does almost as well as the better of H, T, and normally better than D as well. However, S is not E-admissible. It's always *close* to whatever act maximises – something not true of any of the other acts – but it never actually maximises for any **pr**.

Despite being more discriminating, E-admissibility does not seem like an improvement on non-domination. It seems to get things wrong in some intuitive cases, and it violates additional plausible properties for choice rules.

5.6. Improving on non-domination?

Can we improve on $\mathcal{U}$? That is, $\mathcal{U}$ typically contains the reasonable acts, but sometimes contains unreasonable acts too. Can we do better than this? This section considers some attempts to find more discriminating choice functions that still always choose undominated acts. I think, ultimately, we aren't going to be able to do much better than $\mathcal{U}$ as our choice function. This at least rules out the obviously bad acts. There are two ways the following proposals might be read. First we could take them as being on a par with the proposals we have seen so far: that is, as suggestions for how fully rational decisions should be made. I think each will be found wanting on this interpretation. We might instead consider each of these proposals as *somewhat* rational – semi-rational? – ways to determine choice when $\mathcal{U}$ is not discriminating enough. On this understanding, the following proposals' failures are less damning: we have accepted now that fully rational decision making isn't always sufficiently discriminating. We now want to know how to make a choice that isn't rationally determined, and we want to do so while maintaining as much rationality as possible.

I am introducing, here, a sort of *hierarchy of choice rules*. When the fully rational choice rule – $\mathcal{U}$ – fails to be discriminating enough then perhaps other more discriminating but less rational choice rules take over and help narrow down the choice. There is a trade-off between rationally determined decision and fully discriminating decision.

5.6.1. Savage again

Note that the $\mathcal{E}$SAVAGE rule never picks a dominated act. It does, however, violate a lot of plausible conditions on reasonable choice. But there is *something* going right in this rule. It seems to get the right answers, even in the cases of Examples 9 and 10. It also satisfies continuity. So it seems to have some advantages over $\mathcal{U}$. And it seems like it is a bad thing to choose an act if it can only do slightly better than some other, and could be a lot worse. That is, the motivation behind $\mathcal{E}$SAVAGE seems to point in the right direction.

Let's try and find a restriction on choice that is in the spirit of $\mathcal{E}$SAVAGE, but that doesn't have all its bad consequences. Consider the following relation:

DEFINITION 5.6.1 $a \succeq_{\mathrm{PWR}} b$ *iff* $\max_{\mathbf{pr}}\{\mathrm{E}_{\mathbf{pr}}(a) - \mathrm{E}_{\mathbf{pr}}(b)\} \geqslant \max_{\mathbf{pr}}\{\mathrm{E}_{\mathbf{pr}}(b) - \mathrm{E}_{\mathbf{pr}}(a)\}$

This is basically the relation that holds between a and b when a's regret is less than b's when the two are compared pairwise. Because the comparison is done

	a	b	c
$\mathrm{E}_{\mathbf{pr}_1}$	10	1	8
$\mathrm{E}_{\mathbf{pr}_2}$	1	9	1
$\mathrm{E}_{\mathbf{pr}_3}$	1	2	4

Table 5.2.: Cyclic Pairwise Regret

just over pairs of acts, it won't violate the various kinds of contraction consistency properties in the same way that $\mathcal{E}$Savage does. We will see later that it does still violate there properties.

Can we demand that $\mathcal{C}$ satisfy this relation? Unfortunately not, since the relation isn't in general acyclic. For example consider acts that satisfy the properties in Table 5.2.[22] This means that $\succeq_{\mathrm{PWR}}$ would impose inconsistent demands on $\mathcal{C}$. In short, $\mathcal{M}_{\succeq_{\mathrm{PWR}}}$ can be empty.

We can see some analogy here between this rule and the decision rule that Mark Colyvan suggests in Colyvan (2008). His decision rule is also a pairwise comparison of acts that relies on differences in their outcomes. However, he is working in a precise framework, so he takes probability weighted differences of utilities over the states. His rule has more or less the same problems as this rule does.

Despite its failings, there is a feeling that this rule is getting something right. Satisfying this relation in Example 10 – betting on whether two tosses of a coin will be the same or different – gets you the right choice of act (S). In example Example 7 (where $\mathcal{E}$Savage failed), pairwise regret points to a consistent choice: whether or not h is included, f is the most preferred act. This is, of course, some way short of showing that a rule based on this relation would satisfy the various expansion and contraction consistency properties that counted against $\mathcal{E}$Savage. Given that $\mathcal{M}_{\succeq_{\mathrm{PWR}}}$ can be empty, it's not immediately clear how to build a decision rule out of this relation. Perhaps we should demand only that when the maximal set is non-empty, that set should be the choice set. In the same way that we want $\mathcal{C}(A) = ID(A)$ whenever $ID(A) \neq \varnothing$, we want $\mathcal{C}(A) = \mathcal{M}_{\succeq_{\mathrm{PWR}}}(A)$ whenever this is nonempty.

Consider the following choice rule.

[22] Given how similar this table looks to tables of Condorcet cycles in social choice, I conjecture that some sort of single-peakedness property suffices to have the pairwise regret relation to be acyclic.

Definition 5.6.2

$$\mathcal{S}(A) = \begin{cases} \mathcal{M}_{\succeq_{\text{PWR}}}(A) \text{ if } \mathcal{M}_{\succeq_{\text{PWR}}}(A) \neq \varnothing \\ \bigcup_{x \in A} \{\mathcal{S}(A \backslash \{x\})\} \text{ otherwise} \end{cases}$$

The maximal set is never empty for two element sets of acts since you can't have cycles on pairs. This guarantees that this recursive definition never yields an empty set. What it effectively says is "if there is no maximal element, try all the ways of removing one act and look for maximal elements there." Since it is cycles that cause there to be no maximal elements, this procedure will look at all the ways you could remove an element from the cycle, and thus induce there to be a maximal element. Every element of a cycle will end up in the choice set.

Theorem 5.6.1 *$\mathcal{S}(A) \neq \varnothing$ for all finite A.*

Proof Let $|A|$ mean the size of the set A. If $|A| = 2$ then $\mathcal{M}_{\succeq_{\text{PWR}}}(A) \neq \varnothing$. This is just because you can't get the sort of "cycle" behaviour on sets that small. For induction, assume $\mathcal{S}(A) \neq \varnothing$ for $|A| = n$. Now let $|B| = n + 1$. The only difficult case is if $\mathcal{M}_{\succeq_{\text{PWR}}}(B) = \varnothing$. Thus $\mathcal{S}(B) = \bigcup_{x \in B} \{\mathcal{S}(B \backslash \{x\})\}$. But $|B \backslash \{x\}| = n$ for all $x \in B$. Thus by our assumption, $\mathcal{S}(B \backslash \{x\}) \neq \varnothing$. The union of nonempty sets is nonempty, so $\mathcal{S}(B) \neq \varnothing$. ■

I have been a little conservative in only aiming to prove the finite case here. I think the proof holds much more generally, and I hope it is clear how one would extend the proof. There may be some awkwardness caused by infinitely long cycles that may require something a little more sophisticated than $\mathcal{S}$ to get the proof through. The idea, at least, should be clear: make a rule where, if you hit a cycle, take the union of the ways you might break the cycle by dropping an element (or possibly a set of elements).

What properties does this rule satisfy? It is a subset of $\mathcal{U}$. To show this, I first need two short lemmas.

Lemma 5.6.1 *If $a \succeq_{\text{Dom}} d$ then $a \succeq_{\text{PWR}} d$.*

Proof Since $a \succeq_{\text{Dom}} d$ we have that $\mathrm{E}_{\mathbf{pr}}(a) - \mathrm{E}_{\mathbf{pr}}(d) \geqslant 0$ for all $\mathbf{pr} \in \mathcal{P}$. Therefore $\mathrm{E}_{\mathbf{pr}}(d) - \mathrm{E}_{\mathbf{pr}}(a) \leqslant 0$. From this it follows that $a \succeq_{\text{PWR}} d$. ■

This shows that the $\succeq_{\text{PWR}}$ relation is more discriminating than the $\succeq_{\text{Dom}}$ relation. Next we need to show we have a *sort of* transitivity.

Lemma 5.6.2 *If $c \succeq_{\text{PWR}} a$ and $a \succeq_{\text{Dom}} d$ then $c \succeq_{\text{PWR}} d$.*

Proof $a \succeq_{\text{Dom}} d$ so $\mathrm{E}_{\mathbf{pr}}(a) \geqslant \mathrm{E}_{\mathbf{pr}}(d)$ for all $\mathbf{pr} \in \mathcal{P}$. Therefore:

$$\begin{aligned}\max\{\mathrm{E}_{\mathbf{pr}}(c) - \mathrm{E}_{\mathbf{pr}}(d)\} &\geqslant \max\{\mathrm{E}_{\mathbf{pr}}(c) - \mathrm{E}_{\mathbf{pr}}(a)\}\\ &\geqslant \max\{\mathrm{E}_{\mathbf{pr}}(a) - \mathrm{E}_{\mathbf{pr}}(c)\}\\ &\geqslant \max\{\mathrm{E}_{\mathbf{pr}}(d) - \mathrm{E}_{\mathbf{pr}}(c)\}\end{aligned}$$

Therefore $c \succeq_{\text{PWR}} d$. ∎

From the above lemmas, we can show the following.

Theorem 5.6.2 *$\mathcal{S}(A) \subseteq \mathcal{U}(A)$ for all $A \subseteq \mathbf{A}$.*

Proof We need to prove that if d is dominated, then $d \notin \mathcal{S}(A)$. Say that d is dominated by a. If $\mathcal{M}_{\succeq_{\text{PWR}}}(A) \neq \varnothing$ then, since $a \succeq_{\text{PWR}} d$ by Lemma 5.6.1, d cannot be in $\mathcal{M}_{\succeq_{\text{PWR}}}(A)$.

We now need to show that if $\mathcal{M}_{\succeq_{\text{PWR}}}(A) = \varnothing$ we still have $d \notin \mathcal{S}(A)$. The only place we need to check is $\mathcal{S}(A \backslash \{a\})$ since in all other subsets the above result guarantees that d is not in $\mathcal{S}(A)$. If $\mathcal{M}_{\succeq_{\text{PWR}}}(A) = \varnothing$ that means, in particular, that there exists some c such that $c \succeq_{\text{PWR}} a$. If no such c existed, a would have been in $\mathcal{M}_{\succeq_{\text{PWR}}}(A)$. c is in $A \backslash \{a\}$ and $c \succeq_{\text{PWR}} d$ by Lemma 5.6.2. Therefore, d is not maximal in $A \backslash \{a\}$. ∎

Unfortunately, this rule violates Contraction Consistency. Consider a cycle $a \succeq_{\text{PWR}} b \succeq_{\text{PWR}} c \succeq_{\text{PWR}} a$. Now consider $\mathcal{S}(\{a\})$ and $\mathcal{S}(\{b, c\})$. The first of these is $\{a\}$ and the second is $\{b\}$. However, $\mathcal{S}(\{a\} \cup \{b, c\}) = \{a, b, c\}$. Despite not being chosen in either smaller set, c is among those chosen in the larger set. This is because the larger set contains a cycle and thus the whole cycle is part of the choice set. In the smaller sets, there are no cycles and so only the maximal elements make it into the choice set.

If all the acts available are "linear in $\mathbf{pr}(X)$" for some X, then $\mathcal{M}_{\succeq_{\text{PWR}}}$ is non-empty and the rule satisfies Contraction Consistency. By "linear in $\mathbf{pr}(X)$" I mean $\mathrm{E}_{\mathbf{pr}}(f) = u\,\mathbf{pr}(X) + w$. If the graph of expectation against probability is a straight line. This holds of bets on X. Note that this doesn't hold when we take more complex bets. For example in Example 10, you were betting on a coin's landing the same way twice in a row. This is not linear in $\mathbf{pr}(H)$. It is linear in $\mathbf{pr}(HH \vee TT)$, but the other acts on offer are not linear in this proposition. There may be a more general set of cases where $\mathcal{M}_{\succeq_{\text{PWR}}}$ is guaranteed to be non-empty, perhaps something similar to the idea of "single-peakedness" from social choice theory (see Gaertner 2009, pp. 43–9). In the simplified circumstances of choosing between a set of well behaved bets that are all simple bets on the same proposition, then $\mathcal{S}$ is a good choice rule since it is more discriminating than $\mathcal{U}$ but satisfies many of the same properties. In the more general case however, the possibility of cycles means that the rule does not do so well. Despite this, it does seem to get the right answers in most of the example decision problems I have discussed in this chapter.

This is obviously much less than a cast-iron argument that some choice function in this vicinity is the best. Indeed, I have only made gestures towards even showing that this sort of rule is any good at all. It doesn't show much that it looks like it gets things roughly correct in the examples I have discussed.

There is another practical problem with this rule. Given that it relies on pairwise comparisons between the acts, and each pairwise comparison requires comparing maxima ranging over probability values, it is quite computationally intensive.

5.6.2. Levi's rule

Isaac Levi champions $\mathcal{L}$Wald as the best imprecise decision rule. That is, he argues that you should maximise minimum expectation among acts that can maximise expectation. This rule is $\mathcal{E}$Wald($\mathcal{L}(A)$). He argues that only those acts that have some chance of being the best are worthy of consideration. He argues for the "maximise minimum expectation" part by analogy to what happens in precise cases of equal expectation.

He considers a bet like the following: I offer you a bet where you win £1 if heads comes up, but you lose £1 if tails lands up. Should you accept this bet, or refuse it? Both acts (accept, refuse) have the same expectation – £0 – so how do you choose between them? Levi suggests that in this situation you should maximise minimum expectation. He says that the reason to refuse the bet is:

> not that refusal is better in the sense that it has higher expected utility than accepting the gamble. The options come out equal on this kind of appraisal. Refusing the gamble is "better" however, with respect to the security against loss it furnishes. Levi (1974, p. 411)

He suggests the same reasoning works in the imprecise case. We should use "security" as a tiebreaker.

I have already criticised $\mathcal{L}$ and $\mathcal{L}$Wald inherits many of its problems, while adding some new problems of its own. $\mathcal{L}$Wald violates Contraction Consistency, for example, as Seidenfeld (2004) points out.

Example 11 Consider the choice between f, g and the choice between f, g, h.

f £10 if X, nothing otherwise

g £3 if X, £3 otherwise

h £-1 if X, £8 otherwise

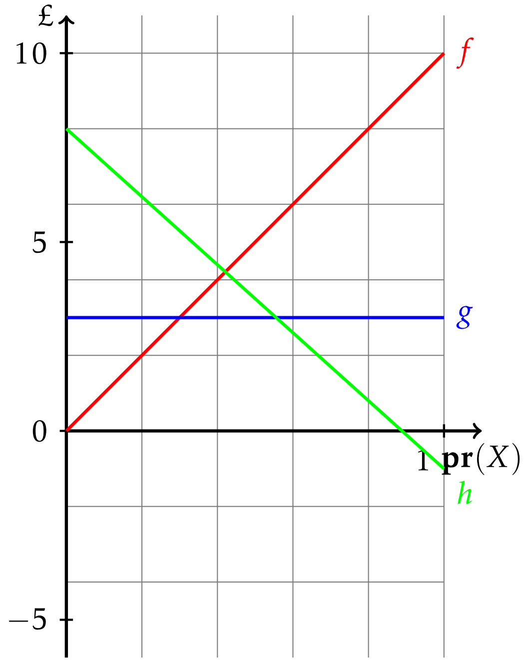

Figure 5.7.: Graph of Example 11

In a choice between f and g, it is g that does best by $\mathcal{L}$Wald. However, adding h means that g is no longer E-admissible and of f and h, f does better. For large values of ρ, $\mathcal{L}$Hurwicz inherits all the same problems.

The trick that causes $\mathcal{L}$Hurwicz to violate these properties is as follows: some act g is E-admissible in virtue of maximising for some value of $\mathbf{pr}(X)$. And it also maximises minimum expectation at some *other* value of $\mathbf{pr}'(X)$ for which g *does not maximise expectation*. So you can make it fail E-admissibility while still having it maximise minimum expectation. Since you have this independence of the two parts of this "lexicographic" rule, you get all the strange behaviour associated with lexicographic orderings.

Now, for ρ small enough – take the case of $\rho = 0$: maximax – you can no longer do this: the places where it wins the Hurwicz criterion test are also the places where it is E-admissible. So the trick that works for $\rho = 1$ is harder to pull off, and in the limit, impossible. I don't know of anyone who argues in favour of maximax-expectation as a choice rule. It rarely accords with intuition in the kinds of examples we have been discussing: typically a modicum of ambiguity aversion seems rational, but maximax is maximally ambiguity-loving. If there is any possible chance that some act *might* be the best, then maximax prefers that act.

What we saw in Example 11 is that g maximises minimum expectation but an act that performs worse on this metric can still make it not E-admissible and therefore not optimal. This means that some other previously non-optimal act must become optimal.

Recall Example 10: where a coin will be tossed twice. $\mathcal{L}$WALD gets things wrong here. Not only does E-admissibility rule out what seems like the best act (S), but then maximising minimum expectation picks out D, which seems like the *worst* act of the four. Likewise in Example 9 – one head versus ten heads plus ε – $\mathcal{L}$WALD results in betting on 10 heads being uniquely rational, for all $\varepsilon > 0$.

I don't think $\mathcal{L}$WALD is a good rule at all. If you are going to violate CONTRACTION CONSISTENCY, it seems like $\mathcal{E}$SAVAGE gets the intuitive examples right while $\mathcal{L}$WALD gets them seriously wrong. If you are committed to CONTRACTION CONSISTENCY, then I'm not sure that any simple decision rule can improve on the relatively permissive $\mathcal{U}$.

5.6.3. Spread-sensitive equivocation

Recall that the objective Bayesian framework endorses three norms, the third of which was Equivocation. I was sympathetic to the spirit of this norm, but I criticised the form it took in Williamson's formulation of the position. Call $\Downarrow\mathcal{P}$ the probability function that is maximally equivocal in $\mathcal{P}$. The details need not concern us here. $\Downarrow\mathcal{P}$ is the probability function in your representor that is the least informative.

We could take maximising expectation with respect to $\Downarrow\mathcal{P}$ to be a decision rule. Obviously, we restrict ourselves to choosing among the nondominated acts, but among those we choose the act that maximises expectation for the maximally equivocal member of your representor. Call this the *Equivocal Expectation*.

DEFINITION 5.6.3 $EE(A) = \arg\max_{\mathcal{U}(A)}\{\mathrm{E}_{\Downarrow\mathcal{P}}(f)\}$

This rule is more discriminating than $\mathcal{U}$, but does it get things right in the examples we have been considering? It gets things right in the "one head versus ten heads plus ε" case (Example 9). However, EE doesn't always satisfy continuity. In fact, we have already seen a case where maximising equivocal expectation seems wrong and this is a case that shows that the rule violates continuity. Recall Example 2 from page 127. In that example you were offered a choice between a sure £0.5 or a bet that wins £$1+\varepsilon$ if a coin of unknown bias lands on different sides in two tosses. The intuitively correct answer is to take the sure money: the expectation of the bet is only better when the chance of heads is very close to 0.5, or rather, the bet looks best only when your credence is very close to a half. But since equivocating means you have credence of exactly 0.5 the bet comes out looking like the best option. Well, you might argue, what we need to do is make the equivocal expectation sensitive to the spread of possible expectations over the whole representor.

Definition 5.6.4 $SP_{\mathcal{P}}(f) = \overline{\mathcal{E}}_{\mathcal{P}}(f) - \underline{\mathcal{E}}_{\mathcal{P}}(f)$

Using this spread penalty we can define the *spread-sensitive equivocal expectation* as follows:

Definition 5.6.5 $SSEE_{\sigma}(A) = \arg\max_{\mathcal{U}(A)}\{\mathrm{E}_{\Downarrow\mathcal{P}}(f) - \sigma SP(f)\}$

The σ reflects your opinion about how much weight to give to the "spread penalty" $SP(f)$.

Unfortunately, this rule also has its problems. Consider Example 10 on page 185 again. This is a similar set-up with bets on coins landing the same way or different ways on two tosses. The spread-sensitive equivocal expectation again chooses the worst act in this example. The problem is that there is another act that does better than it apart from very near 0.5, and this act has the same spread.

The spread-sensitive equivocal expectation induces a complete order on the acts, and the assessment of an act depends only on properties of the act. so it satisfies all the expansion and contraction consistency properties. So in some ways, this choice rule seems better than $\mathcal{L}$Wald, but it is not without its problems.

One might want to extend this idea to a very customisable choice rule: one that has three parameters and that covers the current rule and also $\mathcal{E}$Hurwicz.

$$\mathcal{C}[\alpha,\beta,\gamma](A) = \arg\max_{\mathcal{U}(A)}\{\alpha\,\underline{\mathcal{E}}(f) + \beta\,\overline{\mathcal{E}}(f) + \gamma\,\mathrm{E}_{\Downarrow\mathcal{P}}(f)\}$$

Gärdenfors and Sahlin (1982) interpret Hodges and Lehman (1951) as doing something like this, with γ proportional to the *reliability* of $\Downarrow\mathcal{P}$, $\beta = 0, \alpha = 1-\gamma$. Recall that their model attached a reliability score to each of the probabilities in $\mathcal{P}$. This requires strictly less than what those rules discussed in the next section do. We can recover $\mathcal{E}$Hurwicz by taking $\gamma = 0, \beta = 1-\alpha$. We can recover $SSEE_{\sigma}$ by taking $\gamma = 1, \alpha = \sigma, \beta = -\sigma$. For judicious choices of the parameters, this rule can be made to work for most of the examples we have seen so far. This rule induces a complete order on the acts which means that it satisfies most of the properties we have seen too.

This rule is perhaps too general. With so many adjustable parameters, it can probably be made to rationalise most kinds of choice. For various parameter choices it will fail to satisfy various properties. Consider setting the parameters such that it mimics $\mathcal{U}$WALD: it will now inherit the problems with that rule.

5.6.4. Second-order probabilities to the rescue

Let's return again to the "one heads versus ten heads plus ε" case (Example 9). $\mathcal{U}$ fails to discriminate between the acts in this case, despite the intuitive judgement that there is some reason to prefer the bet on one head. The intuition goes something like this: "Most of your representor thinks that the one head bet is best. Only a small minority think that the ten heads bet is better." This is appealing to some sort of second-order probability: some measure on the representor. This reasoning is unfounded for the same reason that the unwarranted probabilistic assumption in the decision under ignorance case is. You can't really say that "more" of your representor prefer the one head bet. There's no justification for saying that there is more of your representor here or there. The graphs I have shown make it look that way because they suggest that the size of a subset of your representor $\mathcal{P}'$ should be measured by the size of the interval $\mathcal{P}'(H)$. Strictly speaking, this isn't allowed. And even if it were, the size of a subrepresentor would be sensitive to the choice of event to measure it with.

With all that said, there is some intuitive appeal to the idea of measuring the size of the subrepresentor that prefers a over b and using these size comparisons as input into choice. Think of it as the faction within the credal committee that prefer a. Some particular choice of how to measure a representor determines, effectively, a second-order distribution over the representor. One way we might use this to inform decision is to take large finite samples from the distribution and then make the sample credal committee "vote" on which act they like best. This leads to as many decision rules as there are voting procedures and this isn't the place to go into that huge literature.[23] Staying on the intuitive level, we can see that a large majority will vote for one head over ten heads. So this gets things right here. What about the same/different case in Example 10 on page 185? This in fact illustrates an important difference between different voting rules. A majority

[23]But see Gaertner (2009).

prefer Same to Different. But Same is first choice for no committee members (this is just to say that it is not E-admissible). So if the voting rule takes account of first choices only, Same won't be in the running. So if you want to salvage the intuition that Same is the right act, then you need some sort of run-off voting, maybe.

$\mathcal{E}$Hurwicz, the *EE* rule and the current large finite sample from a second-order distribution voting rule all seem to be ways of averaging the opinions of the credal committee members. $\mathcal{E}$Hurwicz averages the committee members' expectations; *EE* averages the committee members' views on the chances; and the current proposal averages their preferences.[24] Each is distinct in its outcomes. The current proposal is, I think, the best of the three; in that it seems it has the potential to get the right answer more often. There are two comments to make on this sort of approach. First, impossibility theorems; second, impossibility of measuring representors.

Impossibility theorems are a big part of the literature on voting.[25] They have the form "No voting procedure can jointly satisfy all of the constraints in *C*", where *C* is some collection of intuitively plausible properties we would like voting rules to have. Will impossibility theorems like this undermine this large finite sample voting method for imprecise decisions? On the current view, not necessarily. One of the critical conditions in Arrow's impossibility theorem – the granddaddy of impossibility theorems – is that the individual preferences that are the input to the preference aggregating procedure are purely ordinal. That is, you cannot use information about *how much more* one committee member prefers *a* to *b* in determining the group preference. This seems reasonable in real voting: are there facts of the matter about interpersonal comparisons of strength of preference? Even if there were, eliciting them would be nigh on impossible. However in the current context, we have been assuming that all the credal committee members share the same utility function, so inter-committee member comparisons of strength of preference should be possible. This fact undermines Arrow's impossibility theorem. Having all committee members have the same utility function was an idealisation that I made, however. It is plausible that we should allow imprecise utility as well. This doesn't however negate the above undermining of Arrow's result. Even if utility is allowed to be imprecise,

[24] Each procedure satisfies something like the properties Ellis (1966) ascribes to averaging functions. They don't exactly satisfy the conditions, but they are close enough that talk of them all being kinds of averaging to be reasonable.

[25] Weatherson (ms.) mentions voting in the context of imprecise decisions, though his remarks are not fully developed.

all the committee members are facets of one person: precisifications of the same person's beliefs and desires. As such, even if the utility functions are different they should be such that the inter-committee member comparisons of utility are possible. Even if different members value things differently, their values are all calibrated to the same yardstick.

The second problem I mentioned above was the impossibility of measuring sets of probabilities. There are some technical problems with measuring sets of probabilities that I have already touched on. One method you might suggest for measuring representors is to measure the interval the representor assigns to some event. The measure will be sensitive to choice of "measuring event". For instance consider $\mathcal{P}$ that consists of all the probability functions that assign $\mathbf{pr}(X) = 0.5$. Y is independent of X and $\mathcal{P}$ is maximally uncertain about Y. Measuring with respect to Y gives $\mathcal{P}$ a size of 1, while measuring with respect to X gives $\mathcal{P}$ size 0. Unless some method can be determined for deciding what is the "right" event to measure the subrepresentors for a given decision, any such method is doomed.

Another method you might consider for measuring sets of probability functions is to consider each function to be a point in an N-dimensional vector space where N is the number of atoms of the algebra. That is, every probability function can be identified with an N-dimensional vector $\vec{\mathbf{pr}}$ where the coordinates sum to 1. For some ordering of the atoms, this vector just lists the values this function assigns to each one. A probability function is determined by its action on the atoms, so every point in the simplex corresponds to exactly one probability function. The set of vectors whose components sum to 1 defines an $N-1$-dimensional simplex. For example, if $N = 3$, the set of coordinates that sum to 1 is a triangle with vertices at $(0,0,1), (0,1,0), (1,0,0)$. Now treat this simplex as a set of points in a vector space and use your standard measure on such spaces to determine the size of sets of probability functions. That is, the space of all probability functions on N atoms is isomorphic to $[0,1]^{N-1}$ and we can use the standard Lebesgue measure on this space to measure the size of representors. This method seems not to have the same sensitivity as the other did, but it will only work for finite algebras, and I can't see how you could justify the choice of measure except to say that it is the standard one. Viewed as just a semi-rational tie-breaker, the arbitrariness of the measure doesn't undermine the method. This method does, however, suffer from language relativity problems. One might replace some atom t but atoms t_1 and t_2 where $t \equiv t_1 \vee t_2$. This would change the measure of representors.

This is just a sketch of a procedure of this sort, but it is up to fans of this sort of approach to fill in the details. I think the strongest thing to say against this sort of approach is that it is monstrously complicated, and for real-world decisions, determining the atoms of the algebra isn't easy.

5.7. The worst rule, apart from all the others?

This brings us back to $\mathcal{U}$. This rule can feel overly permissive, but perhaps this is as far as rationality can get us. When we restrict ourselves to nondominated acts, we have at least ruled out at least some of the unambiguously bad acts.

Echoing Winston Churchill's famous quip about democracy, I think that $\mathcal{U}$ is the worst decision rule, apart from all the others. I suggest that rational decision making can only go so far in cases of severe uncertainty. As I said when discussing decision under ignorance, what gives us the right to expect a determinate answer to all these decision problems? I'd argue that the "half-answer" that nondomination gives us is about as good as it gets as far as rationality determining choice goes in these cases.

If most of the tiebreakers discussed in the previous section end up opting for the same act, then perhaps that suggests that that act is good in a number of ways. Perhaps we could value a certain "robustness to choice methodology". But this will only determine an answer in some situations.

So we return to the other question I started this chapter with. How ought we make decisions? Rationality has only got us halfway; what is going to get us the rest of the way? Perhaps there is a hierarchy of choice rules. At the top we have nondomination. If nondomination is not discriminating enough, we have to move down a level to some sort of tiebreaker rule. This move would have to involve acknowledging that the choice the tiebreaker determines isn't fully rational: I have argued that full rationality doesn't always determine choice, so some element of arationality must creep in. Not all arational tiebreakers are created equal and perhaps the desirable properties of choice rules that we have seen above offer some methods to determine which tiebreakers are less bad. I would argue that such an analysis would put the maximin tiebreaker in Levi's $\mathcal{L}$WALD rule some way from the top. I think there will be a contextual component to which tiebreakers to use, and the last chapter of this dissertation will offer one analysis of kinds of tiebreaker that might be applicable to climate decisions.

6. Scientific uncertainty

> The trouble with weather forecasting is that it's right too often for us to ignore it and wrong too often for us to rely on it.
>
> *(Patrick Young)*

Up until now the discussion of uncertainty has been operating at a fairly high level of abstraction. This chapter and the next focus in on a particular example of uncertainty and decision making. The aim is to show that uncertainty in science is, indeed, multifaceted and diverse; and thus that imprecise probability models are needed since precise probabilities don't do justice to the nature of the scientific uncertainty. The next chapter discusses decision making in this context.

We have already seen some of the ways people have tried to characterise different kinds of uncertainty in section 1.1. What I aim to do differently with this chapter is discuss kinds of uncertainty with a view to what coping strategies we adopt to deal with them. Lo and Mueller go some way towards this, but they treat statistical techniques only. My own perspective will be broader both in the kinds of uncertainty I countenance and in the uncertainty mitigation techniques I consider.

This chapter and the next are more focused, less abstract than the rest of the dissertation up until now. I won't be directly discussing imprecise probabilities all that much any more. But some of the same general lessons and principles carry over from the previous discussion. I take my weaker understanding of rationality to be important in this context: we shouldn't expect the scientific evidence to always determine some precise degree of rational belief. What we believe on the basis of the scientific evidence should be sensitive to the quality and amount of the evidence, and to the kinds of uncertainties that evidence is subject to. My main aim here is not to advocate the adoption of imprecise probabilities for representing climate evidence,[1] but what motivates my advocating imprecise probabilities in

[1] Although, I am in favour of this, too.

the abstract case also motivates what I have to say about climate evidence: be sensitive to the nature of the uncertainties. As I said in the introduction, this chapter and the one following it are motivated by the same concerns that led me to my preferred formal theory of belief and decision: severe uncertainty precludes the satisfaction of the strong standard of rationality that requires determinate answers to all decision theoretic questions; severe uncertainty shows that there is a need to represent the nature and quality of the evidence in some way that makes that relevant to the decision maker.

6.1. Two motivating quotes

How do error and uncertainty enter science? What can we do about it? This chapter explores these questions in the context of scientific modelling, specifically climate modelling.[2] While my main interest is in models of the climate, I think a lot of what I say will translate into other disciplines. I want to start this chapter with two quotes that highlight two aspects of the modelling process.

6.1.1. Laplace's demon: ideal scientist

In trying to characterise what determinism was, Laplace described his famous "demon" who was a kind of ideal scientist.

> We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes. The perfection that the human mind has been able to give to astronomy affords but a feeble outline of such an intelligence. Discoveries in mechanics and geometry, coupled with those in universal gravitation, have brought the mind within reach of comprehending in the same analytical formula the past and the future state of the system of the world. All of the mind's efforts in the search for truth tend to approximate

[2] A version of this chapter appears as Bradley (2011).

> the intelligence we have just imagined, although it will forever remain infinitely remote from such an intelligence. Laplace (1951 [1816])[3]

This demon knows everything there is to know: for him, there is no uncertainty. We of course fall far short of Laplace's ideal scientist, and our actual science is full of uncertainties of various kinds. My aim here is to characterise what kinds of uncertainty arise in science, and point out some of the ways we cope with them.

Laplace attributed three important capacities to his demon.[4] First it must know "all the forces that set nature in motion": it must be using the same equations as Nature is. Second, it must have perfect knowledge of the initial conditions: "all positions of which nature is composed". Laplace makes reference only to positions, but let's grant that he was aware his demon would need to know various other particulars of the initial conditions – momentum, charge, spin – let's pretend Laplace was talking about *position in state space*. The third capacity that Laplace grants his demon is that it be "vast enough to submit these data to analysis". This translates roughly as infinite computational power. I will have less to say about this third capacity, at least directly.

Laplace was operating in a purely Newtonian world, without quantum effects, without relativistic worries. In this chapter, I will do the same. On the whole, climate scientists make the assumption that it is safe to model the climate system as a Newtonian, deterministic system. Even in this environment there are many kinds of uncertainty.

6.1.2. On Exactitude in Science

My second motivating quote is pulling in the opposite direction. Laplace is thinking about the ultimate in accurate prediction, this story from Jorge Luis Borges[5] points to the limits inherent in modelling.

> ...In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations,

[3] This is actually a slightly different translation to the one cited, given in Smith (2007).

[4] Smith (2007) characterises Laplace's demon in this way, and I follow his exposition.

[5] The following quote is the story in its entirety.

> who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography.
>
> Borges (1999, p. 325)

The map here is serving as a metaphor for the modelling process. A map that is too big is like a model that is too slow. And the point I want to make is that modelling is *inherently, inescapably* a process of abstraction, idealisation, ignoring of detail. In a way, at the very heart of the whole project of modelling a phenomenon is the need to introduce error: an exact replica (a 1:1 map) of the phenomenon of interest wouldn't tell you anything. This isn't strictly true. If we had a perfect 1:1 replica of the climate we could run various different scenarios on it and explore the system more in the way of a lab experiment. Simulation as a kind of experiment is an idea explored by Petersen (2012) in the context of climate models. Given the time taken to actually run that sort of model, it would not be a useful *predictive* tool. The actual world is an excellent model of itself, and to know how it will evolve, we only need to wait until it has done so. This is not helpful for many of the purposes we would like to use models for. So what follows isn't a criticism of scientific modelling, but rather marvelling at how much we can throw away and still have informative useful models.

The question we always have to ask ourselves when looking at complex models is: "what is the value added by incorporating these extra processes?" Does the extra insight gained from adding in these extra feedbacks balance out making the model slower and more complex? Does the extra detail justify making the map bigger?

I'm not making any claims that what I say is specific to computer simulation. I more or less agree with Frigg and Reiss (2009) that simulation brings out no new philosophical issues over and above those of other kinds of modelling.[6] Simulation does however emphasise different elements of the issues. The use of computers also introduces interesting new possibilities for error which are worth discussing.

[6] I believe it is up to someone who thinks there is a relevant difference to fill in exactly how this distinction – between simulation and non-computational modelling – can be made.

6.1.3. Outline

This chapter explores the topic of scientific uncertainty. The motivation is to show that we shouldn't be satisfied with glib, easy answers to the question "What should we believe about the climate based on this evidence?"

My examples and much of the discussion will be motivated by climate modelling. However, I will keep things on a fairly "conceptual" and idealised level: much of the detail will not reflect actual practices of working scientists. What I am aiming for is to explore the basic sources of errors and the basic processes for mitigating them. So my conclusions will apply to any model-based, predictive, quantitative science.

We are nowhere near Laplace's demon in our capacities. The next two sections explore ways in which we fail to live up to Laplace's dream. First I look at uncertainty in our measurements of initial conditions (section 6.2). Second I look at uncertainty in our models (section 6.3). Then I turn to looking at the coping strategies we have developed to deal with uncertainty (section 6.4), bearing in mind the cautions of Borges' story. In section 6.5 I discuss some more philosophical topics raised by the foregoing sections. Finally, in section 6.6 I look at when probabilistic prediction is reasonable.

6.2. Data gathering

Scientists gather a lot of information about the phenomena under study. These data are uncertain in a number of ways. For example, the data we collect are finite strings of integers, whereas the quantities we measure can presumably be arbitrarily precise (6.2.1). Our measurement instruments might cause the data to be inaccurate (6.2.2). Or worse, we might not be measuring the quantities we thought we were (6.2.3).

6.2.1. Truncated data: imprecision

Let's say we're measuring temperature at a weather station. The thermometer gives a reading of, say 13.45 °C, but in fact the temperature is more like 13.452 934 °C. Our data of the system's current state is *truncated*. This is one type of uncertainty. It is a fairly mild form of uncertainty: the number is right as far as it goes.

One might complain that temperature is a derived concept, and it doesn't really make sense once you start trying to measure the temperature on smaller and smaller scales. I treat this issue at more length in section 6.5.1. But even granting that, our temperature data is still less precise than it *could* be. There are meaningful significant figures that are not captured in the reading.

What are the reasons for truncated data? Often, the differences in value are not important: for most purposes, all of the temperatures that round to 13.45 °C are indistinguishable, so the truncation doesn't matter. However, if we have reason to believe the system of interest behaves chaotically,[7] then even these tiny differences *can* have an effect. This will be important later. The indistinguishability of different temperatures that round off to the same number is context relative. If you're interested in whether or not to wear a jumper, then all temperatures around 13 °C are dealt with the same way. So in this context they are indistinguishable. If you are interested in the temperature for the purposes of long term predictions of some quantity that depends nonlinearly on temperature, then the temperatures might not be indistinguishable in this context.

Second, our measuring devices have only a limited accuracy, so maybe even where the differences do matter, we might not be able to detect them. Our capacity to store data is also limited, so even if we could actually make arbitrarily fine measurements, we wouldn't be able to store all that data.

Third, the longer the strings of data we collect and feed into our models, the longer the calculations take. So given our limited computational capabilities, even if we could collect more fine-grained data, maybe we would truncate it before feeding it into the computer, in the interests of computational tractability. If predicting Thursday's weather takes until Friday, the forecast is useless. There is a trade-off between precision of the data fed into the model (and the precision of the model itself) and how long that model takes to run. This is Borges' argument again. This also should remind us of this quote from Keynes that bears repeating:

> The long run is a misleading guide to current affairs. In the long run we are all dead. Economists set themselves too easy, too useless a task if in tempestuous seasons they can only tell us that when the storm is past the ocean is flat again. Keynes (1923, p. 80)

[7] This awkward locution is necessary, since it is effectively *unknowable* whether some physical system is mathematically chaotic or not. Indeed, a physical system's being chaotic isn't even well defined: chaos is a property of mathematical equations.

6.2.2. Noisy data: inaccuracy

Sometimes our measurements are not perfect. To stick with the weather example, imagine that the thermometer registered 13.47 °C rather than 13.45 °C. This isn't just a case of truncation: this is an error. This isn't imprecision: this is inaccuracy. This is to assume, of course, that there is a fact of the matter about what the value *should have* been that is different from what it was. I defer a discussion of the metaphysical assumptions underlying my project until section 6.5.2.

There are many reasons why this sort of error might creep in. For example, the equipment might be faulty, or slightly miscalibrated. The more accurate we attempt to make our predictions, the more significant figures we try to determine, the more danger there is that noise – random fluctuations in the system – might overwhelm the signal. For most purposes these errors will be small. But again, if we're worried about non-linear systems, then these small differences can lead to big differences somewhere down the line.

6.2.3. Deeper errors

Measuring physical quantities is a tricky business. It is not a straightforward pre-scientific enterprise that is, somehow, a *precursor* to scientific enquiry. Working out how to measure certain quantities is a long, involved process that makes use of theory and evolves in tandem with it. Chang (2004) charts the fraught history of thermometry, for instance.[8] This suggests another way our measuring of the initial conditions can go wrong.

Imagine if we did not know that there is a connection between temperature and atmospheric pressure. We would not then be controlling for differences in atmospheric pressure, and this would cause our measurements of temperature to be faulty. We would falsely attribute changes in measurement reading to changes in temperature, where actually the change is due to change in pressure. That is, we would be implicitly making an assumption about our thermometer readings not being sensitive to air pressure. If this implicit assumption turns out to be important (and importantly false[9]) then our readings would not be readings of temperature, but readings of some combination of temperature and pressure. To put it another way, we wouldn't be measuring what we thought we were. There is nothing wrong with the measurement *per se*, but with our understanding of what

[8] See also Ellis (1966, Chapter 6)

[9] If the air pressure where we are measuring doesn't change, then this assumption – strictly speaking still false – would not cause misleading results.

we are getting from the measurement.

Given the indirect way we make a lot of measurements in climate science, this is a live issue. What exactly studies of ice cores and tree rings show us about the paleoclimate is a contested issue, so this is another example of where our measurement techniques are not unequivocal (see Schiermeier 2010). How one calibrates size of tree rings with estimates of the temperature at the time that ring was growing involves theory, and a great deal of knowledge of different areas of science. For example, if we made faulty assumptions about what constituted good growing conditions for trees in the past, the size of tree rings wouldn't be telling us what we took them to be telling us.

Indeed, we don't have as many actual measurements of data as we would like. We mostly only have readings for surface-level land values. We don't have as good information about what's going on out at sea or up in the atmosphere (Parker 2006, p. 353). A technique known as "assimilation" is used to "generate missing data" using a particular type of climate model (Talagrand 1997). Errors in these models will lead to faulty "data" being used. This is another way that gathering data can be indirect. Note also that in this way, even our *data* can be subject to worries about model error (which we shall meet in the next section).

To summarise, even at the level of simple data gathering, we must take the data with a pinch of salt. That's not to say that we should be sceptical of all data: there is certainly much information to be gleaned. The point is merely to stress that we can't take an uncritical, naïve view of measurement. Our measurement techniques might not be measuring what we thought they were. This could be due to some physical relationship among the quantities that we are not modelling, or perhaps due to some artefact of the measuring process.

6.3. Model building

Now that we have gathered our data, what do we do with it? How do we turn our data into predictions? We will typically have data in the form of *time series*.[10] A time series is a series of measurements of a quantity (for example temperature) indexed by when the measurement was taken. We will have time series of several quantities – temperature, pressure, precipitation, wind speed – from a variety of locations. This gives us a rough summary of the climactic conditions at those

[10] Other kinds of data aren't all that different. I focus on time series since talking about "evolution in time" is easy and intuitive.

times, in those locations.

Using a variety of statistical tools – regression analysis, radial basis functions, autocorrelation functions and so on – we can explore how these variables change in time, and in response to changes in each other. We can also use our knowledge of the basic physics to inform how we think the relationships among the variables *should* behave. For example, using the Navier-Stokes equation,[11] the basic equation of fluid mechanics, we can guess at how wind speed in adjacent cells – in nearby locations – should evolve in response to each other over time. Lloyd (2009) claims that this sort of support from basic science gives us confidence in our models. I return to this point later.

There are a variety of simplifying assumptions that need to be made here, and none of the techniques used is infallible, so the model building process adds its own sources of error. I outline these sources of error now. I will first list the sources of error, then in section 6.4 discuss how we deal with these errors.

What I should make clear before beginning is that I am interested in systems that will often be chaotic or non-linear. This means that even small errors can quickly becoming worrisome. For a more detailed discussion of the particular issues with nonlinear dynamics and predictability, see Smith (2000, 2007); Smith (1998). The upshot of a focus on these kinds of models is that even errors that are very small can have important consequences.

As Figure 6.1 shows, the climate system is tremendously complex. In philosophical discussions of modelling, it is common to consider simple models of simple systems, like a mass on a spring, or a point mass in orbit around another in an otherwise empty space. It is important to stress that these models are extremely simple. There doesn't seem to be any need to wonder about the process of building the model: it's obvious that it should represent the target system in the relevant ways. In the case of the climate system, finding the appropriate model is hard work.

[11] That said, our understanding of the Navier-Stokes equation is not perfect: indeed, we can't even be assured of the existence and uniqueness of solutions (Fefferman n.d.)

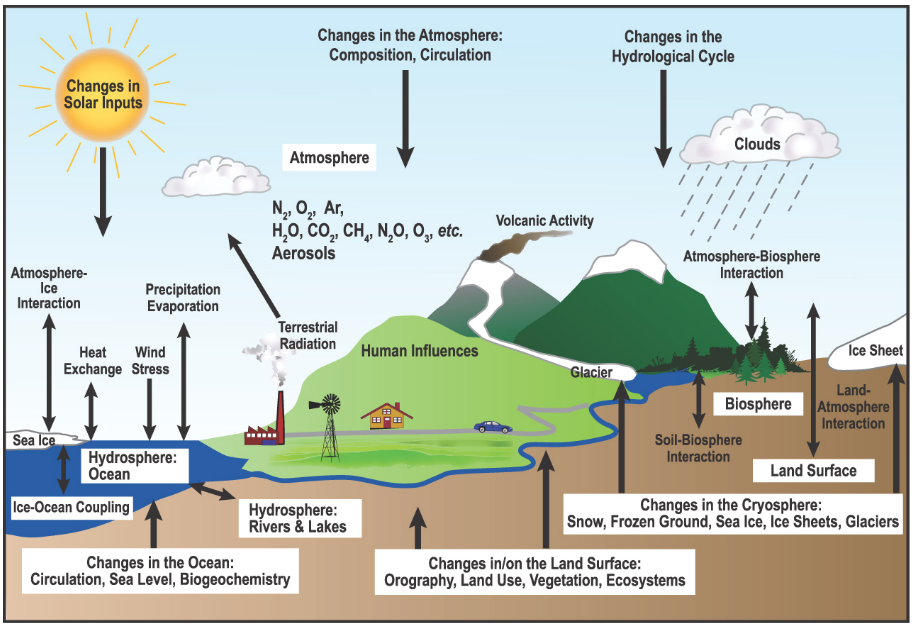

Figure 6.1.: The climate system is tremendously complex

6.3.1. A toy example

To make the discussion more vivid, consider the following toy example.[12] Laplace's demon is trying to predict the evolution of the fish population in my pond. That is, evolution in the sense of "change in population through time"; not as in "evolution by natural selection". He has observed the past evolution of the system and has arrived at a model of the system. He knows that the population density of fish in my pond in a week's time N_{t+1} depends on the current population density N_t. These numbers are normalised such that the absolute maximum population of fish possible is given by 1. He also knows that the future population is adversely affected by how crowded the pond is (captured by the $1 - N_t$ factor). The fish population dynamics are given by:

$$N_{t+1} = 4N_t(1 - N_t) \quad \text{(Log)}$$

This is the famous "logistic map", probably the simplest dynamics that produce chaotic behaviour. May (1976) discusses this model in detail. Figure 6.2 shows how nearby initial conditions diverge radically after a few time steps. This is a figure of 9 initial conditions very close together. The solid line represents the "actual" evolution of the fish population, while the dotted lines represent what the

[12] The example discussed here is due to Lenny Smith and Reason Machete. The "fish population" gloss was suggested by Roman Frigg. The model is discussed in more detail in Frigg et al. (2013b).

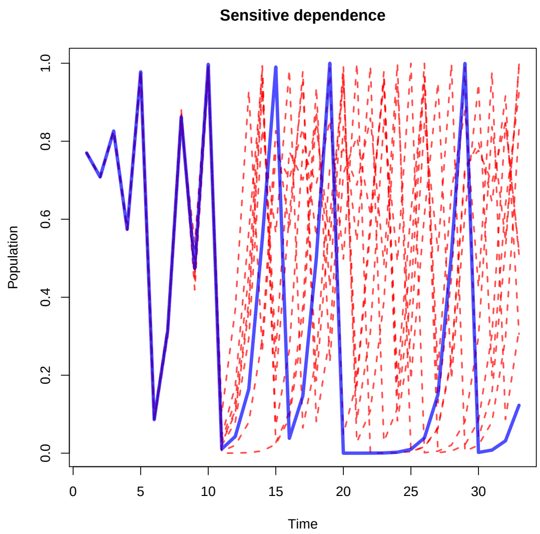

Figure 6.2.: Diverging time series. The blue line is the "actual" initial conditions. The eight dotted lines are nearby initial conditions.

model would predict had you got the initial condition slightly wrong in one of 8 distinct ways. As you can see, after about 12 timesteps, you could be very wrong indeed. This is a characteristic feature of chaotic systems: sensitive dependence on initial conditions. While there is some controversy on exactly how to define what chaos is (Werndl 2009), for my purposes, this won't be important. Many parts of the climate system exhibit this kind of chaotic behaviour. There are many ways in which the logistic map is not a good toy model of the climate, obviously, but they do seem to share this feature that they exhibit sensitive dependence on initial conditions. This sensitive dependence means that sources of very small errors cannot be discounted. This means all the sources of error must be considered and discussed.

6.3.2. Curve fitting

Let's start with an easy problem. We have a one-dimensional time series of data. These might represent daily temperature readings from a particular weather station, or volume of internet traffic through a particular router, or cases of a disease at a particular hospital, or weekly estimates of the population of fish in my garden pond.[13] The question is: given the data, what kind of process could generate that time series? What sort of equation describes the dynamics of how

[13] I have generated a time series of the word count for my thesis for example. The trend is broadly upwards, but all too slowly!

the system evolves?

For any finite data set, any number of different graphs pass through all the points. This is a well known problem with fitting curves to data. If there are n data points, then there is a polynomial of degree $n-1$ that will go through all of these points exactly.[14] And in fact, there are infinitely many polynomials of higher degree that go through each of these points. On what grounds can you say that the process generating the data is represented by *this* rather than *that* function?

Choosing on the grounds of simplicity might be (at least pragmatically) appealing, but there are problems. First, what counts as a simple equation? For example, consider the following equation:

$$\alpha \sin(\beta x) \tag{Fit}$$

For any finite data set, there are values of α and β such that the graph of (Fit) passes arbitrarily close to all the points. This equation is, arguably extremely simple: it has only two parameters. On another measure, however, it's a rather complicated equation, since the graphs it gives rise to are normally "unnecessarily wiggly" given the amount of data it needs to fit. That is, one feels like the curve should bend *only if* it is bending towards a data point it would otherwise miss. A second worry is the following: what guarantee do we have that simplicity is an indicator of truth of a functional relationship? This is a very general worry about science: how do we know that our theory is getting things right about the world?

6.3.3. Structure error

The standard practice when trying to fit a curve to some data is to start with a parametrised family of functions. We can then work out which member of this family most closely fits the data: we can work out which parameters fit the data best. But who's to know that that first step – picking a family of functions – is valid? What assumptions can underwrite that sort of choice? Sometimes we have basic physics evidence that suggests the functional form is, say linear. But other times, we don't have this kind of information. When considering the climate system, for example, there are any number of feedbacks all of them interacting with each other; so we don't really know what model structure is appropriate.

What interpretation can we give to this process? Fitting a curve from a family of functions would make sense if we knew that that data had been generated by

[14] The only case where this is not possible is if two points took the same x value, which is impossible in a time series.

some process that shares that functional form. That is, if we knew that data were generated by some function in that parametrised family, then this would be the obvious way to approach that problem. But it is almost never the case that we are justified in making that kind of assumption. So what does the choice of family mean? What justifies it?

This sort of worry about structure error arises from trying to solve the above uncertainty (6.3.2). I discuss this concern at more length once I have outlined what strategies are employed to overcome those worries.

6.3.4. Missing physics

There are processes in the climate that we know we aren't modelling. We have improved greatly (for example, see Figure 6.3) but there are still many aspects of the climate system left out of our models. In the early 90s, we weren't modelling ocean currents: convection in the ocean wasn't in the model. So we in fact *don't* want our models to fit too closely to the data, because they aren't modelling all the things that are going on in the world.

There are still processes that aren't making it into our models. Some of these processes will have a negligible effect on the climate, but we can't be sure which. Given the nonlinear character of our models, tiny differences due to apparently negligible processes could cause the models to differ significantly from the target system on the sort of timescales we are interested in. Here is a cautionary episode from recent climate science that illustrates this point.[15] At the time of the last IPCC report (Solomon et al. 2007), there was broad agreement among models that there is a poleward shift in storm tracks for extratropical cyclones in both hemispheres (Section 9.5.3.7). However adding more model resolution to the upper atmosphere led to the model predicting an equatorward shift instead (Scaife et al. 2012). Until recently, scientists had expected the stratosphere and the mesosphere to have little impact on the climate dynamics of the troposphere: there simply isn't enough mass of air up there to have much of an effect down here (or so they thought!). This story illustrates that elements of reality that don't make it into the model – even if you expect them to be insignificant – can have an important impact on the model predictions.

[15]Thanks to Erica Thompson for pointing me to this story.

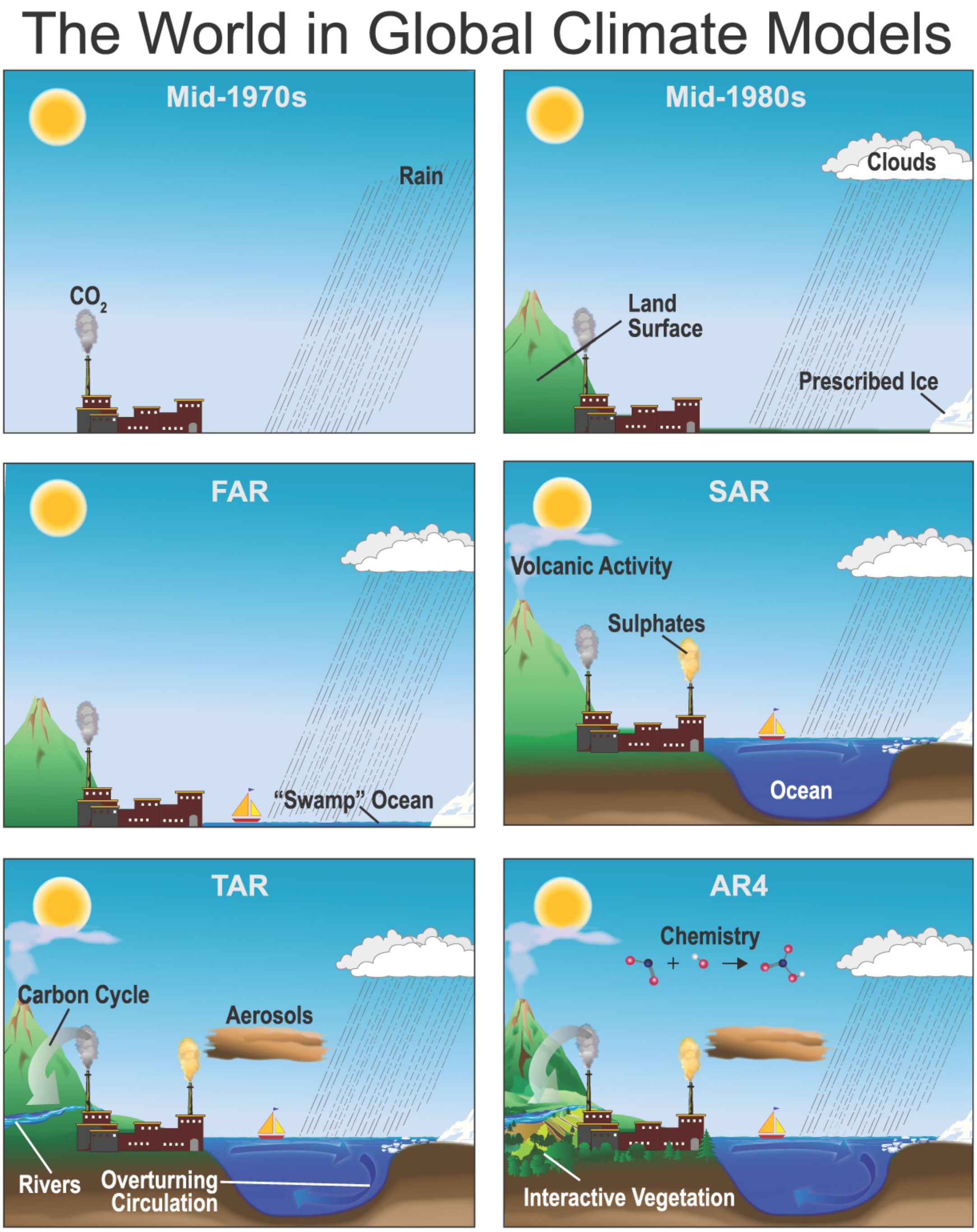

Figure 6.3.: How climate models have improved

6.3.5. Overfitting

As well as structure error, we can be uncertain of the parameters we fix. If there is noise or inaccuracy in the data, then there is a danger of "overfitting". Using Lagrange interpolation to find the polynomial of degree $n-1$ that fits your n points exactly is obviously a crazy idea if you *know* that some of those points are inaccurate! All that you could be doing here is "fitting the noise" which clearly leads to a functional relationship that has very little relation to the underlying process. This is discussed in Hitchcock and Sober (2004). In their own words:

> The data D are bound to contain a certain amount of noise in addition to the information they carry about the underlying relationship between [the variables]. By constructing a relatively complex curve that exactly fits [all the data, the scientist] is bound to *overfit* the data. That is, she is bound to propose a theory which is too sensitive to idiosyncrasies in the data set D that are unlikely to recur in further samples drawn from the same underlying distribution.
>
> Hitchcock and Sober (2004, p. 11)

So the question that needs to be asked whenever there is a choice between a complex, better fitted model and a simpler model with less good fit, is: does the extra model complexity really afford an increase in *predictive* accuracy, rather than just an increase in capacity to accommodate past data? This is the question implicit in the Borges story. One might want to split this question up into two parts. Is the complex model better at interpolating unseen data?; is the complex model better at extrapolating? The first of these asks is the complex model better at predicting where unseen data points will fall in the time series interval covered by the training data? The second asks is the complex model better at predicting where data outside the time frame of the training set will fall? What is the value added of the more complex model? Answering this question depends on how good the models are at prediction, but we don't really have that information for climate modelling. Or rather, we can test how well the model does at interpolation, but we are really more interested in extrapolation: prediction of future data.

This is a problem that worries climate scientists. Obviously, they want to build the best models they can: they want to build models based on the best, richest data available. They want the models to fit the data: if the models didn't fit the data, they wouldn't be very good models! But this data will inevitably contain noise. So how do they avoid the problem of overfitting? I defer answering this question until section 6.4.5.

6.3.6. Discretisation

Another issue is that most of the parameters and quantities we are interested in in atmospheric science are more or less continuous.[16] However, the data we have are discrete: hourly values of particular measurements at particular locations, for example. Quantities of interest vary continuously in time and space. We only have particular measurements from particular places. The computer implementations of our simulations are also discrete: computers are finite state machines, they can't really "do" continuity. So we have spatial and temporal discretisation problems.

This problem leads to "rounding errors". Normally, these aren't all that problematic, but for iterated procedures, chaotic dynamics and the like, these small errors can compound themselves and quickly become serious.

This raises issues about how to go about accurately representing the continuous system discretely, and how to evolve a discrete system so as to track the evolution of the continuous target system.

In fact, there is a deeper problem here. Continuous and discrete systems of roughly the same dynamics can have very different overall properties.[17] The logistic map (which I discussed in section 6.3.1) is chaotic, but the logistic equation, the "continuous version" is non-chaotic, and in fact fairly tame. This "continuous version" is given by the differential equation $\frac{d}{dt}x(t) = 4x(t)(1-x(t))$. The similarity should be clear. Figure 6.4 demonstrates the quite different behaviour of the logistic equation – sometimes called the logistic function – as compared to the discrete logistic map discussed earlier.

Indeed, this must be the case, since one dimensional continuous flow cannot be "chaotic" in any sense, and the logistic equation is one-dimensional. This is a consequence of the Poincaré–Bendixon theorem (Smith 1998, p. 8). Discrete dynamics, however, can exhibit chaos in one dimension, as the logistic map demonstrates.

6.3.7. Model resolution

There is some minimum resolution to the simulation: the world is split into grid squares 100 km on a side, say. See Figure 6.5 for a taste of what Europe looks like

[16] "More or less" because of the concerns about temperature and the like being derived quantities built up out of discrete entities, but see section 6.5.1

[17] Thanks to Charlotte Werndl for pointing me to this issue.

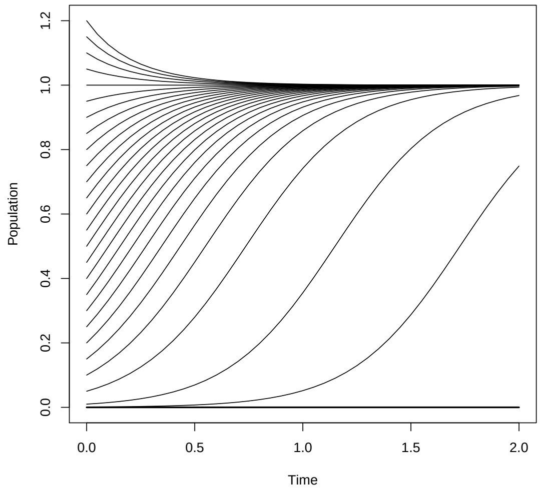

Figure 6.4.: Graphs of the logistic equation: all initial conditions tend to 1.

at this resolution. Processes that happen at smaller scales aren't in the model. For example, clouds are an important factor in climate models. However, typical climate models will have too low a resolution to "resolve" clouds properly, so the effects of clouds have to be put in "by hand". How this is done will be discussed later. Discretisation and model resolution are obviously linked. A particular discretisation method brings with it a particular scale, a particular resolution.

6.3.8. Implementation

Computer programs – which is what simulations are, after all – will inevitably contain bugs. Typical General Circulation Models (GCM) run to hundreds of thousands of lines of code. It would be surprising if there were no mistakes at all in all that code. For example, perhaps a particular function takes "temperature" as a variable, but it wants a relative temperature. If it's used in some module where it is fed an absolute temperature, then this will lead to problems. That is, there are two equally valid ways one might write a function to work out the temperature of a cell at time t_2 given various details of t_1. One approach would have the function output the temperature at time t_2 (an absolute temperature); the other approach would have the function output the temperature *change* between t_1 and t_2 (a relative temperature change). Each is perfectly valid, but it should be obvious that things will go seriously wrong if some process were expecting to receive an absolute temperature and got a relative one or vice versa. The process would just receive a number, and would manipulate it as normal, but the mistake

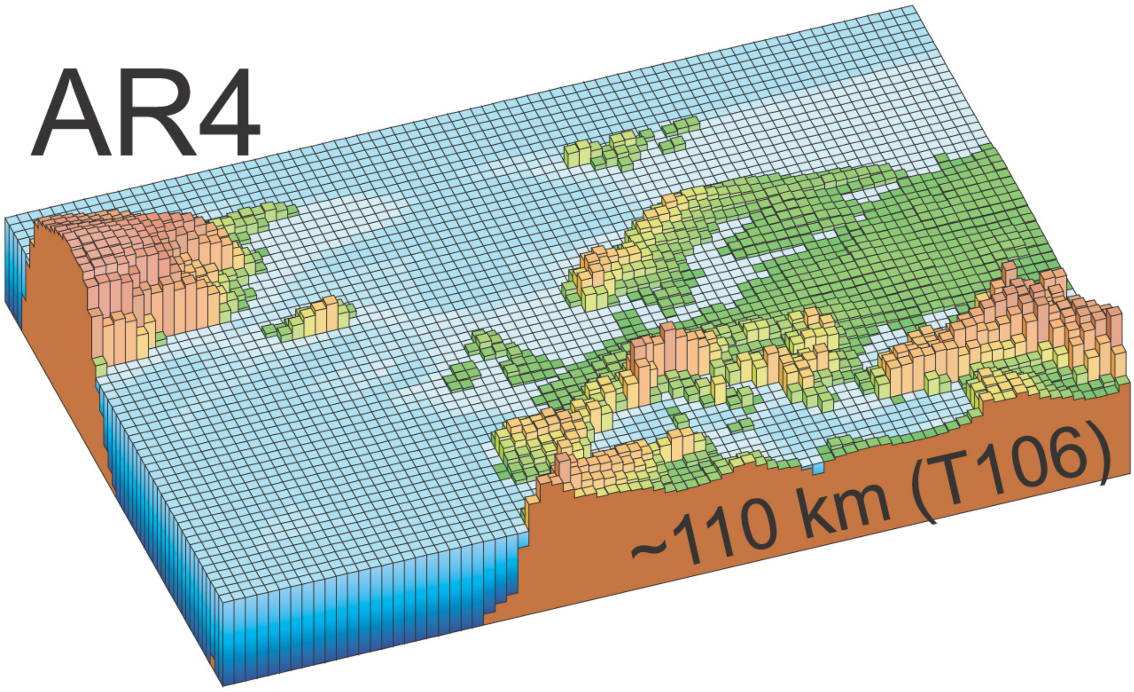

Figure 6.5.: Europe as viewed by the models in IPCC AR4 (2007)

would propagate through subsequent processes and invalidate the results.

At an even lower level, hardware can be faulty or buggy (Muldoon 2007). For example, the Pentium FDIV floating point unit bug caused some calculations to go wrong by as much as 61 parts in a million. Hardly a big deal in general, but if you are dealing with potentially chaotic systems, these things can cause noticeable errors (after a few iterations). Recall Figure 6.2: nearby initial conditions diverge. Different chip architectures implement arithmetic differently, so running the exact same program on different chips can lead to different results, even without hardware faults. Even different compilers, compiling the same code on the same chipset can lead to differences in output.[18] So this is an extra source of problems.

On the other hand, computers are finite state machines. They can't really "do" chaos, since there is a limit to the accuracy with which they can represent numerical values. As Smith (2007) points out, if you run a computer simulation of a chaotic system for long enough, it will settle into a cycle. This is related to the inability of digital computers to properly represent continuity.

[18]Lenny Smith in conversation

6.4. Coping with uncertainty

This has so far seemed to be a fairly negative discussion. However in this section, I want to outline why things aren't as bad as they might be. I want to highlight some ways we have of maintaining confidence in our models in spite of all the above sources of error.

Before starting I want to point out that these ways of maintaining confidence come in two varieties. One variety are the methodological ways of maintaining confidence. These are methods we can use to cope with uncertainty. For example, derivation from theory (section 6.4.2) and ensemble forecasting (section 6.4.4) are such methods.

The other category is harder to name. These aren't *methods* we can follow, but rather, aspects of the output that give us confidence. Robustness (section 6.4.6) and past success (section 6.4.7) are of this sort. These properties of output aren't really things we can aim for, but they are things that warrant confidence if they turn out to hold.

6.4.1. Make better measurements!

An obvious response to at least some of the issues raised above is that we need to make better measurements. This is a reasonable point and there are certainly cases where better measurements could help, but it certainly won't solve all our problems. While better measurements might mitigate against some of the problems, it won't really *solve* any of them. Better measurements would give us more significant figures in our data, but wouldn't really stop it being the case that the data are truncated.

While mitigation of error is obviously still a good thing – and it's normally all we can do – I want to point out that it doesn't solve the problems. For practical purposes, good mitigation is as good as solving the problem. How good the mitigation has to be depends on the practical purpose in question. How many significant figures are needed such that the remaining error in initial conditions doesn't lead to an unacceptable level of error in the output depends on what counts as "acceptable" error in the output. Parker (2009) discusses the notion of "adequacy for purpose". The point is that it is important to be clear what you want out of your model, since this affects what you can put in it and still be confident of the results.

There's a question of whether better measurements are a cost-effective way of increasing adequacy. For example, we could put thousands and thousands of buoys in the seas to get more detailed readings of temperature and so on across the ocean. Would this extra data (gathered at considerable cost) lead to better models? Would other uses of that money have led to *even better* models?

The idea of making better measurements relates to Karl Popper's idea of "accountable" predictions (Popper 1982). John Earman discusses "Popper's demon" as a weakening of Laplace's demon who is limited to only use finitely precise initial conditions (Earman 1986). This demon can, however, calculate how accurate its predictions are given the initial data. The idea is, for any given desired precision of output, the demon knows how accurate its initial conditions would need to be. So, given the model the demon has, it knows how quickly nearby initial conditions diverge. From this, it can work out how accurate its initial data have to be in order for the predictions to be adequately precise. Given the complexity of climate models, even this ideal is unachievable for us mere mortals. That is, in order for the trajectories to have not spread out too much yet we would need data orders of magnitude better than our current data. And of course, model error – which we shall discuss later – undermines the adequacy of the prediction even if we could collect that data.[19] This idea is related to those discussed in section 6.4.3.

It's important to be clear that for our purposes, we can be confident that making better and better measurements would be uneconomical, even if it were possible. Other sources of confidence in model output must be sought.

6.4.2. Derivation from theory

One thing that helps with confidence is to know that our models and our parametrisations have a sound basis in basic physics. This is something Lloyd (2009) mentions as a source of "confirmation" for climate models.[20] If a model bases its heat transport processes on basic thermal physics principles, then that is good reason to think it will predict well. It's clearly not sufficient to guarantee good predictions that a model be based on theory, but it is a good start!

Here is an example of how a simple model of the climate can be derived from some basic physics. McGuffie and Henderson-Sellers (2005, p. 82 ff) derive a simple 0-dimensional "energy balance model" (EBM) from basic physics with the

[19] Actually working out the precision of the data required is a non-trivial task, computationally speaking. But we are operating on a conceptual level, so we won't worry about this.

[20] It seems Lloyd is using "confirmation" in a non-standard way: to mean something like "giving us confidence".

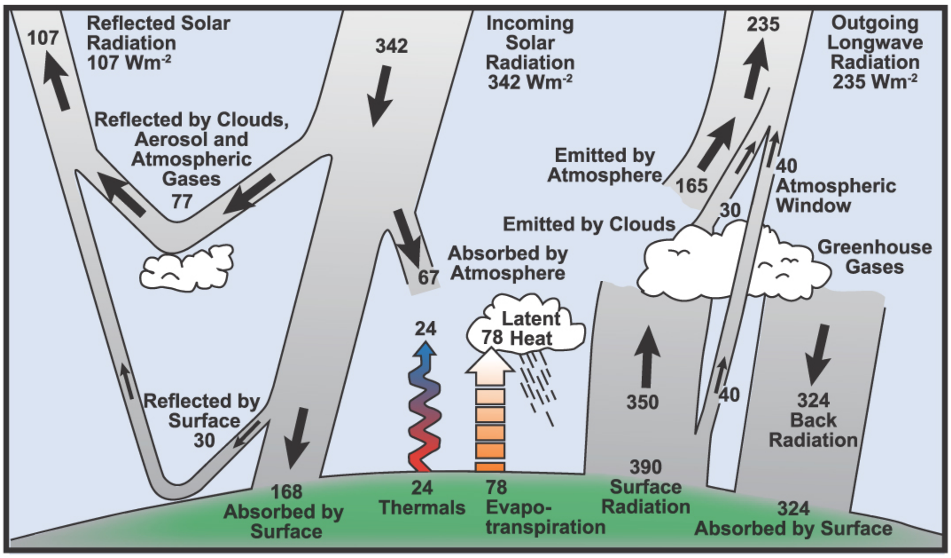

Figure 6.6.: Basic energy balance of the Earth

help of some geometry. The basic processes involved in these simple models are summarised in Figure 6.6.

Imagine a lump of matter subjected to a constant source of heat energy. If the lump did not emit any energy, its heat must increase without bound. Now imagine that the lump – which we will call a planet – emits radiation (loses energy) at a rate that increases with temperature. The temperature of the planet will increase until the outgoing radiation balances with the incoming radiation (Houghton 2009; Pierrehumbert 2011).

From thermodynamics we know that on average, the energy input and the energy output of the Earth must balance.[21] The Sun provides effectively the only external source of energy. How much sunlight the Earth's surface reflects (its albedo) will be an important factor in this process, as will how much of the energy radiated from the Earth gets through the atmosphere. The "solar constant" S is a measure of the amount of energy per unit area that the Sun provides. The amount of energy received by the Earth is $\pi R^2 S$ where R is the radius of the earth. That is, the energy received is the energy per unit area provided by the Sun times the area of the Earth "visible" to the sun at a time. This will be a circle of radius R. A certain amount of this energy is reflected back by the surface and atmosphere.

[21] In reality, given the uptake of heat by the deep ocean and similar processes, there is a genuine question about whether this claim is true for any given timescale (Dave Stainforth in conversation). This is, however, the basic assumption that underlies this derivation.

The albedo (α) is a measure of the fraction of energy reflected back in this manner. So the energy absorbed by the Earth is:

$$(1-\alpha)\pi R^2 S \tag{6.1}$$

Imagine a "black body" that absorbs all incident radiation and whose only outlet for energy is emission of radiation. The spectrum of wavelengths and amount of radiation emitted depends only on the heat of the body. The Stefan–Boltzmann law states that the energy emitted per unit area is proportional to T^4, where T is the temperature of the body. The Earth emits heat over its whole surface of $4\pi R^2$. So we can calculate the energy emitted by the Earth's surface per unit time to be:

$$4\pi R^2 \sigma T^4 \tag{6.2}$$

Where σ is the Stefan-Boltzmann constant and T is the temperature of the Earth. The energy input and output must balance. Putting these terms equal to each other and cancelling some terms we get:

$$(1-\alpha)S/4 = \sigma T^4 \tag{Temp}$$

Using $S \approx 1370\,\mathrm{W\,m^{-2}}$ and $\alpha \approx 0.3$ and solving for T gives you a value of about 255 K which is about −18 °C. If there were no atmosphere, this temperature – the effective temperature – would be the surface temperature. This "effective temperature" in fact differs from the surface temperature due to the effect of greenhouse gases (GHGs), which act to insulate the surface.

The surface temperature and the effective temperature will differ depending on how efficiently the atmosphere absorbs the radiation emitted from the Earth. From (Temp) and an estimate of this efficiency, we can calculate what the temperature of the Earth is, and how it would change if the atmosphere changes.

Greenhouse gases act like a "one way mirror" for radiation: they let in solar radiation (mostly UV and visible spectrum wave lengths) but they absorb and reflect some of the outgoing radiation from the Earth (mostly infrared spectrum). That energy is absorbed back into the Earth and thus drives up the surface temperature. The surface temperature and the effective temperature will differ depending on how efficiently the atmosphere absorbs and reflects the radiation emitted from the Earth. This "greenhouse increment" – the effect on surface temperature due to energy absorption by the atmosphere – is about 33 K at present. Plugging these numbers in gives us a surface temperature of 288 K, or almost 15 °C. Increasing GHGs in the atmosphere will increase the amount of radiation absorbed and thus will increase the surface temperature.

From some well confirmed physics we have derived a simple model of the climate system. More complex models require more work, but the same general pattern can be seen in them as well. The next stage would be to build a 1-dimensional climate model that captures the fact that more solar radiation falls on the equator than falls on the poles. We would then need to model how heat flows from hotter regions to colder regions. This model-building process would be informed by the well-understood basic physics of heat transfer (McGuffie and Henderson-Sellers 2005, pp. 85–86).

Knowing that the model we have built rests on these physical theories gives support to the model. The model could still be wrong: some of the abstractions, idealisations or approximations might be inapplicable given our purposes. But on the whole, at a suitable level of abstraction we can take confidence from this model building process. Note that this source of confidence relies on taking the basic physics for granted. If we were being so sceptical as to question the fundamental theories we appealed to, then no such confidence would accrue from this procedure. Or rather, only *conditional* confidence would accrue: *if the basic physics is correct* then the model is on solid foundations.

6.4.3. Interval predictions

Returning to our fish population model; if the demon knows the exact initial conditions – current population – he can predict the future population of fish with perfect accuracy. But consider a different level of description. The demon is now trying to predict the future population based, not on an exact initial condition, but on some *set* of initial conditions, some "patch" of the initial state space. This might correspond to some measurement of initial conditions with finite precision (truncated data). In the logistic map example this would be making predictions given some small interval of initial conditions. Recall Popper's demon.

Obviously the demon cannot give pinpoint predictions for this sort of initial condition, since different initial points in the initial condition patch end up at different places once subjected to the dynamics. The demon can, however tell you exactly what the patch of the final phase space is that houses all and only those initial conditions that started in that initial patch.[22]

The problem with this "patch prediction" is that for something like the logistic map, it quickly becomes uninformative (see Figure 6.7). Because of the chaotic

[22]Recall this demon has unlimited computational capacity: it can complete the supertask of putting every point in the uncountable set through the dynamics

nature of the logistic map, the initial error grows rather quickly, so after a few iterations the interval you predict is rather big, and it soon covers the whole interval.[23] This is obviously not a very helpful prediction. But it is interesting as a measure of the predictability. That is, if the interval blows up quickly, this indicates that the system is hard to predict. This relates to the discussion of accountability above. For more on accountability and its relation to interval forecasting and ensemble forecasting see Smith (1996).

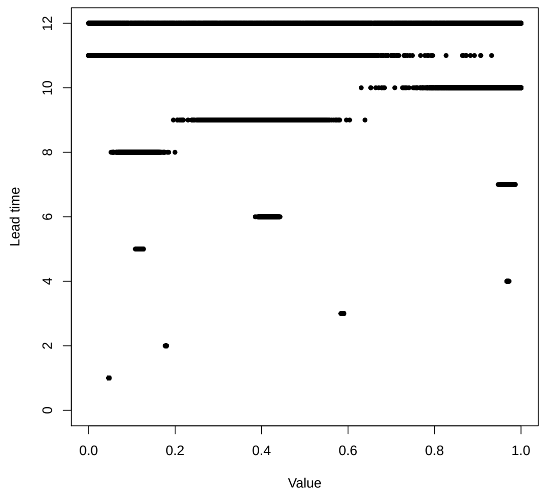

Figure 6.7.: Graph of interval predictions. The interval covered at each lead time (after each number of iterations).

Patch prediction can help against imprecision (section 6.2.1) but not much else. This "exploring the uncertainty" is an important idea in climate science. While not directly useful in improving the precision of our predictions, this is a good way of learning about our models. We learn what interactions lead to big changes by seeing which small differences in parameters have big effects.

This is all taking place at a fairly abstract level. How the comments in this section (and those following it) relate to actual practice is complicated. But my claim is that something roughly like these procedures underlies the (much more complex) practices of working scientists. What scientists actually do will be far more sophisticated (and in some ways more constricted), but the basic idea of exploring the uncertainty through ensembles is the same.

[23]Strictly speaking it will become dense over the whole interval.

6.4.4. Ensemble forecasting

Say instead of a patch of initial conditions, we give the demon some probability distribution over the possible initial conditions. For example, a "noise distribution" peaked around an observed value. The demon could evolve this distribution through the dynamics and tell you what the distribution of final conditions looks like. Now, what if the demon weren't quite vast enough to submit all these data to analysis? There are, after all, uncountably many initial conditions in any patch or distribution, to subject to the dynamics. And each initial condition is of infinite precision. So that's an awful lot of computation to do. So, how would a demon of unreasonably large, but still finite computational resources fare? What the demon might do is sample from the noise distribution, say 1024 initial conditions of finite precision. That is, he would take his measured initial conditions, and take some probability distribution centred on the measured values. This distribution accounts for his understanding that the measured initial conditions could be slightly wrong. He then samples from this distribution to give him his ensemble of initial conditions. He could then subject each of these initial conditions to the dynamics, and take the distribution of model outcomes as representing the distribution of possible system outcomes.[24] What the demon does is pick out a set of possible initial conditions – sets of possible values for current temperature, precipitation, wind speed and so on – and put each of these possible descriptions of the current climate through the model dynamics. The demon can then explore what future evolutions of the system are possible in the model. Figure 6.8 shows an example of an ensemble forecast of the logistic map at various lead times. As you can see, the error spreads out fairly quickly.[25]

Why is ensemble forecasting better than "best-guess" forecasting? Before ensembles, the standard practice in weather modelling was to take the best available data and the best available model and to run one simulation in as much detail as possible. Now, we sample from a noise distribution and run each ensemble member through some model. If there are 1024 ensemble members, then this takes 1024 times as much computing time as one model run, so we need to use a simpler model than if we just had the one member. So how does this lead to better outcomes? What is the value added by the more complex ensembles approach?

[24] A common practice is to take some ensemble of initial conditions and adopt a uniform prior probability over them. This is more or less the same as the sort of procedure I have been describing.

[25] These graphs are generated by the same data as the previous figure.

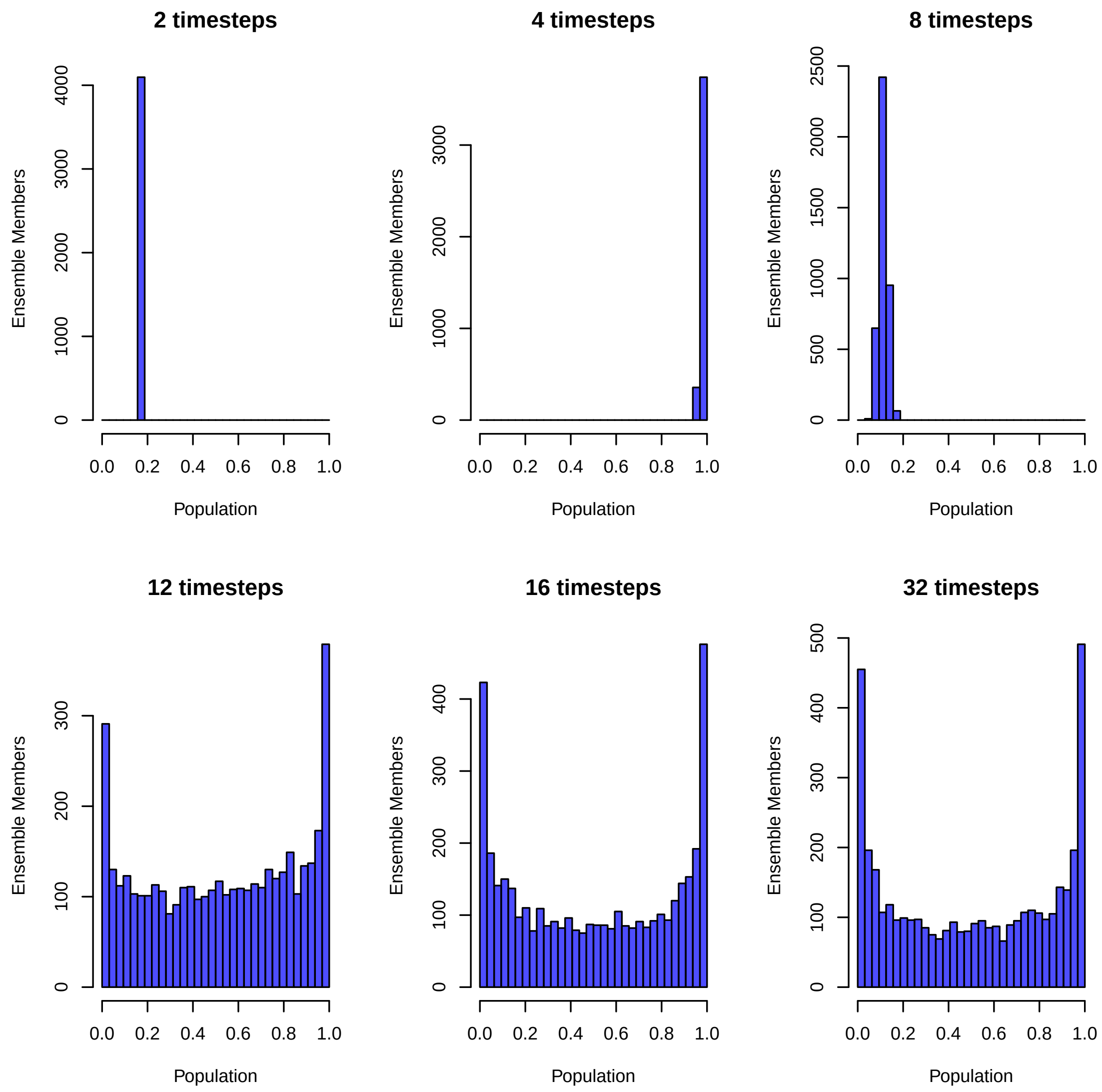

Figure 6.8.: An ensemble of outputs at various lead times for the fish population. The distribution of ensemble outputs after various numbers of iterations.

The problem with "best-guess" forecasting is that initial conditions very close to the best guess end up looking very different (recall Figure 6.2). Loosely speaking, the weather exhibits some kind of chaotic behaviour: sensitive dependence on initial conditions. So the ensemble method shows you what possibilities there are. Smith (2007) discusses how ensemble forecasting predicted a serious storm that hit England in 1990 which would have been missed by best-guess forecasting (see pp. 10–16,139–143).

The "patch prediction" I discussed above also has this feature of highlighting the space of possible evolutions, but ensemble forecasting goes beyond this in "quantifying the uncertainty" by means of the output ensemble. In this way, ensembles can help quantify our knowledge of noise, and so it can deal with inaccuracy (section 6.2.2), as well as imprecision.

One can make an ensemble of initial conditions but it is also possible to sample from a distribution on a *parameter value*: this "uncertainty distribution" accounts for possible errors in parameter values in the same way the noise distribution does for initial conditions. In this way ensemble modelling can help to deal with some worries about model error, as well as initial condition uncertainty. However, we are still working within a parametrised family of functions so similar worries apply as have been already voiced in section 6.3.3. What guarantees do we have that we are getting the model structure right? Initial Condition Ensembles – ICEs – and Perturbed Physics Ensembles – PPEs – are often used together to try to sample as much of the uncertainty of the model structure as possible. We can take the brute statistics of the ensemble members, or we can use sophisticated methods of weighting the ensemble members by how good they are (Tebaldi and Knutti 2007). One method is to score models on how well the can predict twentieth century climate. This folds in aspects of the next method I will discuss. ICEs and PPEs are joined by a third kind of ensemble: Multi Model Ensembles – MMEs – are formed by taking the outputs of several different research groups, using different model structures, and looking at the statistics of these grand ensembles. This does something to mitigate the worries about being tied to a particular model structure, but not much. There are only a handful of groups running large scale genuinely distinct GCMs. So we are not really properly sampling the space of possible models. Indeed, models share structure and code, so even the small sample we have isn't really a good sample (Tebaldi and Knutti 2007, pp.2067–8). It is an "ensemble of opportunity".

PPEs and ICEs investigate uncertainty about numerical values, while

> MMEs investigate uncertainty about model structure. For the former, it is relatively straightforward to conceptualize, and then sample systematically within, the space of plausible alternatives; for the latter it is not. Parker (2010b, p. 989)

Even when we are sampling a parameter value, there is a worry that accidental features of the set-up may lead to bias. Different ways of describing what is ultimately the same parametrisation can lead to a uniform sampling of this quantity being different. Stainforth et al. (2007a, p. 2154) give the example of "ice residence time in clouds" versus "ice fall rate in clouds". These are not linearly related so uniformly sampling one leads to a non-uniform sampling of the other. This is a nice real-world case of language relativity. They are obviously describing the same physical thing, but the choice of one over the other leads to differences in sampling.

Given that ensembles don't deal with all the kinds of error, it is a mistake to take ensemble distributions as being probabilities of events in the world. But there is certainly decision-relevant information to be gleaned from these ensembles, and more information than just the possible spread of outcomes; but probably less information than a full probability distribution function over possible outcomes. How much information, and exactly what information there is in ensemble outputs is a question still under discussion (Frigg et al. 2013a,b; Parker 2010a,b; Stainforth et al. 2007a; Tebaldi and Knutti 2007).

> The frequency distributions across the ensemble of models may be valuable information for model development, but there is no reason to expect these distributions to relate to the probability of real-world behaviour. One might... argue for such a relation if the models were empirically adequate, but given nonlinear models with large systematic errors under current conditions, no connection has been even remotely established for relating the distribution of model states under altered conditions to decision-relevant probability distributions.
> Stainforth et al. (2007a, p. 2154)

Questions remain about what interpretation can be given to this ensemble. The world has some particular initial condition, and there is some determinate outcome that will happen, so what sort of probability is the ensemble distribution? This is not a new problem: a similar issue has been extensively discussed in philosophical literature about statistical mechanics (Frigg 2008a; Uffink 2007).

There are however differences between interpreting statistical mechanical probabilities and climate projection probabilities. It is standard to interpret ensembles of simulation outputs as probabilistic evidence. That is, we interpret the ensemble of outputs as giving us information about the probability distribution function (PDF) of possible future events in the target system. Recall that the ensemble is discrete, while the PDF in the target system is presumably continuous. So do we treat the output ensemble “as if” it is a representative sample from the system PDF and infer things about the system PDF from it? Since the model PDF doesn’t account for all the sources of error we have discussed, this seems unwarranted: how do we know that model distribution is not importantly different from the “system distribution”? It seems a more subtle approach to model output is needed. Stainforth et al. (2007a) criticise ensemble PDFs for not being adequate for decision support. Frigg et al. (2013a,b) demonstrate how such an approach can go wrong.

6.4.5. Training and evaluation

As we’ve already seen (in section 6.3.7), some processes that happen at scales smaller than the grid scale are important. How are they dealt with? Let’s take the example of clouds. Clouds affect the albedo of the Earth. Clouds also have an insulating effect like greenhouse gases do. These two effects act against one another: increasing albedo cools the earth below the cloud layer, insulation warms it. Which of these factors wins out – the direction of the net effect – depends on a number of factors related to the make-up of the cloud. The intensity of these effects differs for different kinds of cloud.

Clouds can’t be put directly into the model, because they are smaller than the size of the grid square. So we have to add correcting factors to the albedo and insulation parameters for the square. The exact values of these parameters will depend on other variables that affect clouds – e.g. humidity, temperature – and will vary as these vary.

We aren’t really simulating clouds, merely simulating their effects. We aren’t calculating what particular values of humidity and so on will generate what shape clouds, but rather directly what effects these clouds would have on albedo and the like.

How do we work out what values these parameters should have? Derivation from theory is more or less impossible, since these aren’t parameters that appear in basic physics theories.[26] We do this by "training" the model on some data. So for example we see what values the parameters should have in order to correctly predict climate data from 1970 to 1990. We then test these parameter values by seeing whether the model can successfully predict climate data from 1990 to 2010. If this is successful, that gives us confidence that the parameter values are successfully compensating for the unsimulated cloud effects.[27] This kind of statistical inference is a fairly widespread practice. Hitchcock and Sober (2004) discuss *why* it is a good practice: in short, it is a good balance between having the model fit the data and avoiding overfitting.

This helps us mitigate against worries about curve fitting (section 6.3.2) and overfitting (section 6.3.5). And to some extent, training gives us confidence that we've avoided structure error (section 6.3.3). That is, if the model predicts "unseen" observations, then that seems to suggest it is getting something right about the structure of the world.

There is something strange about this process, however. We take some parametrised family of models and we use a statistical technique to find a model with good fit. The statistical technique is "blind" to the physical interpretation of the model. That is, we're trying to parametrise clouds, but the statistical technique doesn't "care" about whether the parameter value it spits out is at all physically meaningful. One would have thought that a reasoned process that took account of the physical meaning of the parameters (and their interaction) would easily outperform such blind fitting.[28] So how do we get predictive models out of this process? And what does this mean for our attitude to the models? Can we interpret the parameter value as telling us something about the world?

As Hitchcock and Sober point out, the motivation for the training and evaluation methodology is "purely instrumental in character." They continue: "Even if we know that the true curve is a polynomial of degree r, it may well be that the curve of degree r that best fits the noisy data one has at hand will fare worse in future predictions than a curve of lower degree." (Hitchcock and Sober 2004, p. 14).

Petersen (2000) criticises this blind statistical procedure, calling it "bad empiricism", in contrast to "good empiricism" which involves having your parametrisations be informed by basic science. Petersen wants us to derive our "cloudiness" parameter from out understanding of the basic physics of cloud dynamics. This is

[26]That said, we can get some information about the possible form they might take from theory.

[27]The actual practice is significantly more complicated than this, but the basic idea is the same: fit parameters with some data, validate it on other data.

[28]Katie Steele pointed out how odd this process is.

an admirable aim, but is it a practicable one?

One might argue that Petersen is right that in the limit, we should try and have all parts of our model be physically motivated. But we need projections of the future evolution of the climate *today* so we had better make them with our current best physics. Our current best knowledge doesn't include a capacity to have physically meaningful parametrisations for all processes we need to model, so if we want to model those processes, we had better be bad empiricists. Bad empiricism is better than no empiricism. Is failing to model a particular process better than modelling it in an ad hoc way? Covey (2000) thinks not and I agree.

If the choice is between not modelling clouds (implicitly assuming they have no overall effect on the climate) and modelling clouds in an ad hoc "bad empiricist" way, then bad empiricism seems the better choice. But is this a false dichotomy? Would it be better to somehow learn to live with only getting the kinds of predictions we can justifiably secure with good empiricism alone? Going down this line of reasoning might leave us with rather few predictions. This "theoretical limit" would involve resolving clouds and all other currently parametrised processes, and looks rather like if we were to follow this line to its end, we would be falling into the trap that Borges outlined. So it's not even clear that bad empiricism is bad in the limit. That said, there is certainly something to the intuition that the more we derive from well confirmed physical theories, the better we will do. So it seems that "avoid ad hoc parametrisations" does serve as some sort of regulative ideal. As Parker (2006) discusses, there is a commitment to model the physical processes in "the right way" as much as possible. Ad hoc fudges that just "save the phenomena" are a last resort: physically meaningful, realistic solutions are much preferred. So it seems there is a sense in which we do take ourselves to be trying to "get at" the equations Nature uses.

What about the worries I raised earlier about simplicity being no guide to the truth of a functional relationship? In fact, Akaike (1973) shows that a measure of predictive accuracy explicitly includes a "simplicity measure" in terms of the number of adjustable parameters. Akaike defines a measure of predictive accuracy that can be broken down into two components: one that measures the function's success so far (how close it is to fitting the training data); and another component that is effectively the number of free parameters (a crude measure of simplicity). Hitchcock and Sober (2004, section 5) discuss this surprising result in the context of model-building.[29]

[29] See also Sober (2002).

There is a worry that future climate data won't be like past climate data in the way we would want it to be for this source of confidence to work. We are training models on their ability to model the climate when GHG levels are at or below current levels. GHG concentrations are rising, so the future climate system will be unlike the past in having these higher levels of these gases.

> Statements about future climate relate to a never before experienced state of the system; thus, it is impossible to either calibrate the model for the forecast regime of interest or confirm the usefulness of the forecasting process. Stainforth et al. (2007a, p. 2147)

This goes some way to undermining the use of training and validation. There is still some confidence to be gleaned from these procedures since we might think that models that do well at predicting past data will also be better at predicting future data, just as long as the difference in GHG concentrations isn't too large.

Another use that past data is put to is it is used to weight the different ensemble members when producing ensemble statistics. That is, the predictions of those models that better predict past data count for more in generating a PDF based on the ensemble. For the reasons just discussed, ability to predict the past needn't entail ability to predict the future.

A further worry is that different parameter settings can be equally good.

> [M]ultiple locations in parameter space can be identified which yield simulations of present climate of comparable quality to the standard...while predicting a range of transient climate responses.
> Murphy et al. (2007, p. 2000)

If we take successful prediction to be confidence-warranting because it suggests we have got something right about the parameter values, then this underdetermination is worrisome. If different settings of the parameters can lead to equally good prediction, then which parameter setting truly reflects the physical target system? Different errors may cancel out (Oreskes, Shrader-Frechette, and Belitz 1994). The next section turns this worry on its head and suggests that it may in fact be a good thing.

6.4.6. Robustness

We don't just have the one model: different research groups have different models, built using different data sets and techniques, for different purposes, at different resolutions, incorporating different mechanisms. Agreement between these disparate models is an important source of confidence in their predictions (Lloyd 2009; Parker 2006, 2009). Even a single model can exhibit a kind of robustness if the same predictions come up for different settings of the parameters.

My usage of the word robustness is fairly broad and covers concepts that others might want to distinguish: for example agreement among models and insensitivity to changes in parameter values. I believe there is enough in common between these various concepts that they can be discussed together, and I choose to use the word "robustness" to refer to them. In short, I allow "robustness" to refer to agreement across ICEs, PPEs and MMEs.

Let's first consider parameter robustness. This might also be called "insensitivity". The idea is that if some particular prediction is insensitive to the value of a particular parameter, then this prediction is robust in this sense. There are some caveats to this characterisation of parameter robustness. If a result is robust across the range $[7, 10]$ for some parameter x, and we believe the actual value of x to be near 1, then this robustness is not confidence-warranting. This might seem like a trivial point, but it's not one I've seen discussed anywhere. Or again, even if we have a robust prediction for parameter values $(1, 5]$ but pathologically weird behaviour below that, then we still don't have that much confidence in the prediction.

The idea is that robustness is confidence-warranting if we have robustness *for a range of values around the physically expected value of the parameter*. And this important caveat also points to the reason behind the confidence: if we have that robustness, then wherever the actual parameter value is in that expected range, the prediction of the outcome variable will hold.

Imagine we knew we had the right model, but we were missing a parameter value. Here, robustness of a result around the possible values of the parameter does warrant confidence. Wherever the parameter is in the possible range, the outcome is the same, so since we've got the model structure right, we know the outcome is assured. To put it the other way, if a result is not robust across different values of an unknown parameter, then we have no reason to be confident in our prediction of that effect based on some estimate of the unknown parameter.

I take this to be the kind of thing people have in mind when thinking about robustness of simple models. We have a simple model motivated by basic physics and an unknown parameter. Say we're investigating heat flow in some liquid for which we don't have a good estimate of thermal conductivity. Some effect that is robust across different values of this unknown parameter is likely to be correct. We think that our model is getting something right about the structure of the world. So robustness in the model suggests that, counterfactually, whatever nearby value Nature could have picked for the parameter, the result would be the same. So this gives us confidence that the result will hold in the actual world. This is similar to the sort of "robustness analysis" that Weisberg (2006) discusses. The details of this line of thought will depend on what position one takes on the theory-world relation.

Introducing model error into this discussion adds another level of complexity. Now consider robustness of a result across different climate models. There are elements of the model that we *know* to be unphysical. There are parts of the simulation that we know aren't mapping onto structure of the world (or are mapping onto system-structure only very indirectly). So how is robustness of the model evidence that the result is assured in reality?

If different models with different idealisations make the same prediction, that suggests that the prediction is not an artefact of this or that idealisation (Muldoon 2007). What we have here is evidence that the prediction is due to the shared structure of the models which we hope is structure they also share with the world. This is defeasible evidence, obviously, but evidence nonetheless. What we have is evidence that there are no troublesome systematic errors in our models.[30] We can think of running different models of the climate as exploring a part of *model space*. Let's imagine there is some model which is a perfect model of the actual world. We can now put our intuitions about robustness like this: as long as we are exploring the part of model space that contains the perfect model, then we can be confident that robustness warrants confidence. However, we can't always be sure of achieving robust predictions. What to do with "discordant evidence" – when our predictions disagree – is an open question (Stegenga 2009).

The missing stratosphere/extra-tropical storm tracks story I mentioned earlier might give you pause for thought here. In 2007, the IPCC authors felt that there was enough consensus to mention model agreement in their report. They did mention a dissenting opinion, but there was nevertheless something of a consensus (Solomon et al. 2007, Section 9.5.3.7). This consensus has now been overthrown. This is an interesting case because the shared error of the earlier models can't be straightforwardly attributed to some shared modelling artefact.

[30]There are, however, almost certainly some systematic errors that have not yet become troublesome...

What the models "shared" was a lack of stratospheric resolution. I can't see how failure to model some physical process could cause the sort of spurious agreement that this episode illustrates. One might worry that agreement with previously existing models is being used as a criterion for adequacy of a model. This sort of sociological interdependence undermines the confidence that robustness can otherwise warrant.[31]

6.4.7. Past success

For something like weather forecasting, we have plenty of evidence that our models are doing pretty well at predicting short term weather. This gives us confidence that the model is accurately representing some aspect of the weather system. Parker (2010a) discusses this issue. Past success at predicting a particular weather phenomenon, together with the assumption that the causal structure hasn't changed overmuch does allow a sort of inductive argument to the future success of that model in predicting that phenomenon. Past success suggests that problems with implementation, hardware, and missing physics are not causing the models to go wrong. If these things were causing the model to go wrong, we wouldn't have had the past success we did, in fact, have.

Climate modellers do not have access to this source of confidence. First, they don't have the same wealth of past successes: they are predicting things that are still in the future, so they don't know if they've predicted accurately yet! Climate models have only so far managed to retrodict already existing data (see section 6.4.5). Weather forecasters are predicting days in advance and they have years of data on their performance. Climate modellers are trying to predict years, decades into the future and only have a couple of decades worth of data on how climate models have done. And even this data isn't very helpful: how useful is it to know how a particular climate model from the 1980s did at predicting the 2000s? We now have much more sophisticated models, but these models have not yet demonstrated any predictive success.

Climate predictions are conditional on particular CO_2 emission scenarios, as well. So we can't even be sure that the causal structure isn't changing in ways that could undermine our ability to predict. Also, if the world is warming, past data and data from the warmer future will not be similar in all the ways they need to

[31] This point has come up in discussion periods at the Bristol Philosophical Issues in Climate Change Workshop (May 26 2011) and the European Philosophy of Science Association conference in Athens (October 6 2011). Thanks to the participants at those events.

be to underpin the material induction. We can't be sure of the "stable background" we need to ground our induction (Norton 2003). So, even if we had past success, we couldn't be sure that the conditions hadn't changed so as to invalidate any confidence this success gave us.

6.4.8. Summary

Table 6.1 summarises how each of the coping strategies discussed fares in relation to each of the sources of error. Imprecision and inaccuracy can be mitigated against by making better measurements and running ensemble forecasts rather than "best-guess" forecasts. Deeper worries about measurement might be offset by agreement among models which suggest that the quantities we are measuring are the right ones. Basic physics might also give us confidence we are on the right track as far as getting the relevant quantities right goes.

Some physical parameters can be measured independently of the climate modelling so better measurements can sometimes help with worries about parameter error. For example, the thermal conductivity of air, or the viscosity of water can be determined independently of the role they play in climate models. Robustness of a prediction across an ensemble of parameter values mitigates against worries about parameter values, even in the absence of such direct techniques. We can also train models on past data to determine unknown parameters.

Error in the structure is somewhat harder to deal with. Finding parameter values through training on past data involves starting with some parametrised family of functions, so we can't train our models to account for error in picking the model class. If basic physics suggests that the structure ought to be roughly *this way*, then that warrants some confidence that we aren't subject to structure error there. If a prediction is robust across different models with different structures, this gives us confidence that the differences in structure do not contribute to making this prediction go wrong. Likewise, past success suggests we have got the model structure roughly right. For similar reasons, past success and robustness give us some confidence that the physics missed out of our models is not leading our models to be seriously wrong. Missing physics can also be accommodated by "training" unphysical parameters that correct for the processes we are leaving out.

That a model trained on some past data can still retrodict unseen data partially mitigates worries about overfitting. Past success can be seen as warranting extra confidence of the same kind. If the model isn't too sensitive to changes in its parameters, this also suggests that the model is not over fitted.

	Better measurements	Theory	Intervals	Ensembles	Training	Robustness	Past Success
Imprecision	M		PM	M			
Inaccuracy	PM		PM	M			
Deeper		PM?				WC?	
Parameter	PH	PH		M	M	PH	PH
Missing physics					PH	WC	WC
Structure		WC				WC	WC
Overfitting					PM	WC	PH
Discretisation						PH	WC
Resolution						PH	PH
Implementation						WC	WC

M: Mitigated — PM: Partially Mitigated — WC: Warrants Confidence — PH: Possibly Helpful

Table 6.1.: Summary of errors and coping strategies

Worries about discretisation and resolution effects can be somewhat allayed by results not being sensitive to changes in resolution, and by past success at modelling the system at that level of resolution. Deep worries about implementation can be dealt with in the same way.

I have outlined a variety of sources of error in scientific modelling. I have also explained some ways we have of maintaining confidence in our predictions despite these errors. I tried to diagnose how these methods allowed us to remain confident in the face of error.

6.5. Philosophical reflections

In various places in the above discussion of uncertainty I glossed over some important philosophical topics. I return to a couple of them now.

6.5.1. A digression on the meaningfulness of derived quantities

Temperature is a shorthand for "mean kinetic energy". So "temperature at a point" is meaningless. Temperature is a quantity defined for some volume, and is defined by the average energy of the particles in that volume. Imagine shrinking the volume of interest until it is of a size with the atoms. Now imagine that at a particular time, that volume is empty.[32] But say at some later time, a particle whizzes through the volume. The mean kinetic energy of the box would suddenly jump from 0 to k where k is the kinetic energy of that particle, and then it would jump instantaneously back to 0 when the particle leaves the volume. On this scale, it should be obvious that temperature is not a useful quantity. In short mean kinetic energy is just not a stable, meaningful, useful concept on scales of the order of the mean free path of the atoms.

A similar meaninglessness is apparent if we consider the millionth decimal place of a temperature in a "normal sized" volume. The millionth decimal place might well change radically as particles enter and leave the volume. So at this resolution, again, the quantity is useless: it is pure noise.

But the above discussion does not undermine temperature as a meaningful, useful quantity *on the appropriate scales*. Of course, at some level, at some degree of precision, the whole concept of temperature breaks down; but we are not near that limit in practice. Even if the thousandth decimal place of a temperature

[32] Let's pretend we've never heard of vacuum energy: we are being good Newtonians…

reading might well be meaningless, there are meaningful significant figures that we aren't measuring. It does not mean that striving for more accuracy is wasted effort: it simply means that there is a limit.

The limit isn't an absolute limit on what science is capable of: if we were working on the tiny volume scale, we simply wouldn't be using quantities like temperature. We would be using the resources of quantum mechanics, rather than classical statistical mechanics and thermodynamics. So while arbitrarily precise temperature readings don't make sense as the ultimate goal of better measurement, it is still the case that there are meaningful significant figures we aren't capturing (again assuming there is a fact of the matter). Those significant figures that are meaningful that we aren't measuring could well be worth knowing about.

6.5.2. Realism and the "True" model of the world

I want to highlight a couple of points about the above discussion that have a bearing on our attitude towards the model. I mean by this the attitude we ought to take vis-à-vis whether the model is adequately representing the target system.

Throughout this chapter I have been talking as if there really is some set of equations that "Nature" uses to evolve the real world system. Our aim is to get our model as close to these equations as possible. Is this true? Does it even make sense? Whatever the answer to those questions, it is certainly the case that it is unreasonable to assume that the "True" equations of the world are in the class of models we are currently exploring. We *know* there are myriad ways our model class *must* fail to contain the True equations if they exist.

So how ought we talk about the practice of modelling? It seems it can only be a verbal shortcut to talk of our getting at the True model: a metaphorical flourish. Indeed, the model and the system are just *very different kinds of things*. I've pointed to some of these differences above, but let's make them clear again. First, the target system – the physical world – is a physical thing. The model is... what? A computer program? A mathematical construct? Some combination of the above? So it's not even clear what it means to say that the model "gets things right" about the world. What does it mean for a model to be "like" its target system?

One might want to retreat to a kind of instrumentalism here and say that the model is "like" the target system in that the things in the model that represent the observables – temperature, wind speed and the like – are "like" those things in the world. That is, one might take the view that the model and the system agree on the level of the observables: on the level of the temperature data and precipitation data; one could then refrain from making any claim about whether the way the simulation manipulates the data bears any resemblance to the physical processes. So temperature data in the model is like temperature in the world. But there is still a concern that these are quite different kinds of things. That this particular part of computer memory registers a value interpreted as "13.45 °C" doesn't straightforwardly translate into a fact about the world. Is comparing data measured in the world with data generated by a model comparing apples to oranges? Perhaps we would like to say that all we are really trying to do when we are modelling is to summarise the actual data as best we can. But then why run simulations into the future?

First, there is a certain sense in which we really *do* take the models as telling us something about the world. These models are supposed to be predictive: they are supposed to tell us how the actual climate will actually vary.[33] So there are certain parts of the model that it seems we really have to be realist about. And this isn't an idle philosophers' realism: decisions costing billions of pounds are made on the basis of these predictions' correctness. How's that for ontological commitment? Recall also the commitment to model processes in "the right way" that Parker (2006) mentions.

However, there are still parts of the model that we know are unphysical: all the sub-gridscale parametrisations are simply ad-hoc ways of keeping predictive accuracy. There's no attempt to model the phenomena, we are merely trying to accommodate the measurements. So taking this "trying to get at the True Model" line too seriously is misleading. We have a commitment to accurate predictions too and sometimes this leads us to approaching things in unphysical ways. This line might bolster my earlier response to Petersen (2000) on "bad empiricism".

6.6. When to take probabilistic predictions seriously

In this section I aim to outline what needs to be the case in order for a probabilistic prediction based on statistical or model evidence to be a reasonable guide to decision. When, in short, are the worries about severe uncertainty of the previous chapters not applicable?

Norton (2003) attempts to ground inductive inference in local facts, as opposed to giving a global, formal theory of induction. When various local facts about

[33] These projections are conditional on a particular emission scenario, but they are still supposed to be saying something conditional about the actual world.

there being some stable background for the induction hold, inductive inference can go ahead. As a theory of induction this just pushes the problem back a level: what could possibly ground the local facts about the background remaining stable in the future? But Norton's theory is helpful in that it emphasises what we must take for granted in order to warrant induction.

Let's take a specific case of statistical inference and see where the assumptions enter. The plan is to start with a discussion of an idealised case, and then progressively remove idealisations to see which results survive which idealisations. We are tossing a coin of unknown bias. After N tosses it has landed heads H_N times and tails T_N times (obviously, $H_N + T_N = N$). As N increases $\frac{H_N}{N}$ becomes a better and better estimate of the objective chance of heads, with probability one.[34] This is so because of the *Law of Large Numbers* (LLN). What assumptions are we making in the above inference? What properties of the set-up does LLN rely on? The LLN holds for *independent and identically distributed* (iid) samples. What this means is that each coin toss has the same chance whatever the previous outcomes were, and each toss has the same chance of heads as the previous tosses. For coin tosses we can take these properties for granted. If we are sampling from a large population, we need to be careful to control for various sorts of statistical bias before we can make this sort of assumption. If you were making decisions that depended on the next coin toss, the value $\frac{H_N}{N}$ is appropriate for use as a decision weight, provided N is large enough. If $N = 1$ then the above fraction will be 0 or 1, and it won't provide good enough evidence to be useful for decision.

Let's move now to the less simple case of weather forecasting. We have several decades of daily weather data, and in some cases we have records stretching back a hundred years or more. Using this data we can estimate the average temperature or rainfall. This isn't, however, a particularly useful calculation. In temperate climes, the effect of the seasons means that the mean temperature isn't a useful piece of information. If you were interested in tomorrow's rainfall in a region that has a monsoon season and a dry season, total annual rainfall divided by 365 is unhelpful. Maybe we have just got the reference class wrong here. If it is July and you are interested in tomorrow's temperature, what you want is the average temperature of past Julys. But even this will typically be too coarse-grained a statistic to be useful. Weather forecasting goes beyond these crude statistics by

[34] I am going to help myself to objective chance talk here. I don't think anything fundamentally different happens when you don't allow yourself this kind of talk, but things get slightly more complicated. See Kreps (1988, Chapter 11) for a treatment of statistical inference in purely subjectivist terms.

measuring weather patterns and using physics-based models to run simulations of future weather. We know roughly how areas of low pressure affect wind speed and direction, we know how areas of cloud will move so we can predict where they will drop their rain. In these ways, we can do better than the brute statistics in predicting weather outcomes. The models that centres like the UK Meteorological Office run are ensemble models (see section 6.4.4). Thus they give probabilistic predictions of future weather.[35] These predictions seem formally similar to those offered in the climate case. I was critical of the possibility of such forecasts for climate, so it will be instructive to see what the distinction is that means that the weather probabilities can be legitimate. Indeed, we have already seen all the parts of the difference: there are certain sources of error that climate models are subject to that weather models aren't, and there are certain sources of confidence that apply to weather models but not to climate models.

Short range weather forecasts – one to five days ahead – don't suffer from the problems with lead times that climate models do.[36] That is, the small errors don't have time to blow up on the timescales that weather forecasters deal with. Weather forecasters can afford to build models with higher spatial and temporal resolution, since they don't have to be run so far into the future. The state of the art in 2007 – when the last IPCC Assessment Report was published – had climate models with a resolution of around 100 km, while the Met Office's Numerical Weather Prediction (NWP) model boasts a resolution of 1.5 km over the UK.[37]

As well as these advantages, weather forecasts benefit from some extra sources of confidence. As well as copious weather data, we have copious data on how well our weather models predicted. This information has helped improve the model and also gives us confidence that the model is generating useful, decision relevant information. We can be confident that weather models are *adequate for purpose*, to use the phrase from Parker (2009). Given that we have only really had big GCMs for a couple of decades, and that the prediction lead times of interest are thirty, forty, fifty years in the future, we don't have the same kind of confidence that the information generated by the current generation of climate models is

[35] `http://www.metoffice.gov.uk/news/in-depth/science-behind-probability-of-precipitation` Retrieved 30 July 2012.

[36] In this respect long range and seasonal weather forecasts are a little more like climate than short-range weather. To make the distinctions clear, I will focus on short range forecasts, but there is a spectrum of forecasts between weather and climate.

[37] `http://www.metoffice.gov.uk/research/modelling-systems/unified-model/weather-forecasting` Retrieved 30 July 2012.

accurate or decision relevant. Ultimately, the same basic physical processes – heat transfer, fluid flows – are responsible for climate processes and weather processes. However, the proximate causes are different and the weather-processes are better understood than the climate processes. For example, we don't know how much heat the deep ocean can absorb. Our better knowledge of the weather processes is due to the wealth of information we have about the weather's behaviour and the timescales involved. To take up the deep ocean again, we don't know its capacity for absorbing heat energy because we don't have enough data on how the deep ocean responded to previous increases in temperature.

On the sort of timescales that weather models deal with, we can take for granted that we have the stable background that we need to ground the right sort of inductions. For climate, the situation is rather different. We have no historical data on how the climate responds to the high levels of CO_2 currently in the atmosphere. We don't know that the response will be similar to the responses that our models predict. For the reasons given above, we have *some* confidence that the model response adequately captures the important features of the system response, but as we saw, climate models haven't passed the same stringent confirmation tests that weather models have. The models are based on the data we do have, and we know that those data are relevantly different from the actual situation. On some more basic level – at the level of the thermodynamics – we know what the effect of the CO_2 should be. At this level, we do have the stable background. But this isn't enough to scale up our predictions to the climate system, with all its unknown feedbacks and nonlinear dependencies. The situation is, in some ways, like the following: I roll a green die several times and record the outcomes. I then give you this information. You use this information to make predictions about what will happen when I roll a blue die. There's a sense in which the data I have given you shouldn't be taken to be decision relevant, given the important differences between the data generating regime (green) and the current decision regime (blue).

I take the above discussion to show what it takes to be confident of your models' probabilistic predictions, and also to show how we don't have that confidence in the climate case. We can of course make probabilistic predictions for climate events: many people do. But the aim of the above discussion was to show how these predictions *should not* be used as a basis for decisions. The kind of epistemological warrant for taking the probabilistic predictions seriously just isn't there in the climate case. We are in a state of severe uncertainty and that should affect how

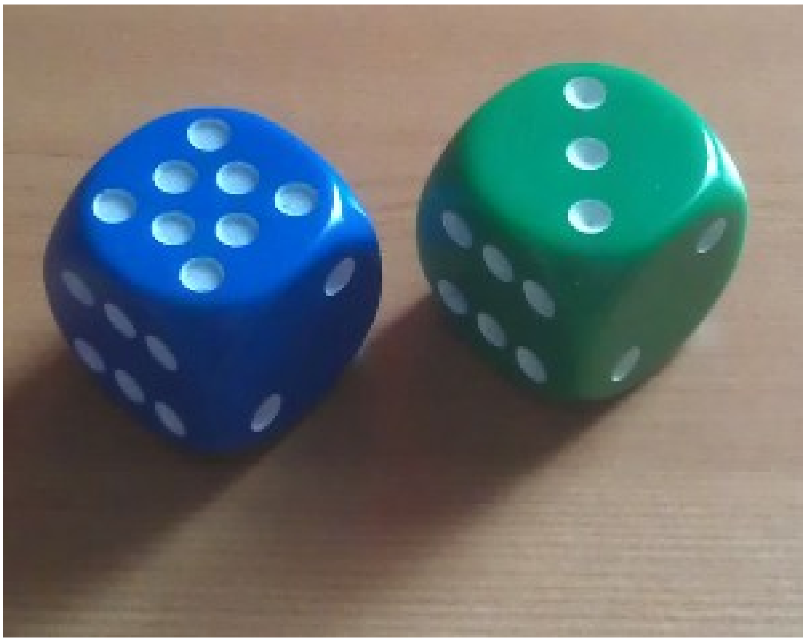

Figure 6.9.: A caution against overstretching your inference: a blue die (l) and a green die (r)

decision making proceeds. Thus the pessimistic decision theoretic conclusions of the previous chapter hold. What climate models can give us already is a "non-discountable climate change envelope": a set of possible futures that we cannot rule out (Stainforth et al. 2007b). Can imprecise probabilities help here? Perhaps we could take each ICE or PPE as providing a "non-discountable PDF", and then we take the set of them as representing the current state of knowledge. This contrasts with the actual current methodology of model-averaging in a number of sophisticated ways to generate a single probability distribution function (Murphy et al. 2007; Tebaldi and Knutti 2007). So instead of using sophisticated model-averaging techniques, we just take each member of the multi-model ensemble as providing a probability in your representor. Decisions would then be made on the basis of this representor. This would focus attention on those parts of the evidence where the models disagree, and on those parts of the evidence where there is a agreement. I take the important point of such an exercise to be the focus on which conclusions are robust, and on which conclusions are not: such a focus is important, and the model averaging techniques are in danger of sweeping such important information under the carpet. Note however that robustness may be due to incomplete or biased sampling of the "model space" rather than due to some feature of the world. So this imprecise probability approach is not a panacea for all model-error worries.

If we take this *non-discountable PDF* approach, we will be stuck with the problem that we ended the last chapter with: there will be circumstances when the imprecise decision rules fail to discriminate among different acts. The evidence will not be good enough to give us advice. The next chapter will discuss what to do when we end up in such a situation.

7. Climate decisions

All models are wrong, but
some are useful.

(George Box)

In the last chapter we saw that climate science has many sources of uncertainty. In previous chapters we have seen that decision making in the presence of such uncertainty can be very difficult. This short chapter will look at what we can do when faced with climate decisions.

The main aim is to argue that only the weaker standard of rationality is applicable given the uncertainties present, and to show what this means. The aim is not to argue directly for the use of imprecise probabilities in climate decisions. I will argue that just because we don't have a decision relevant probability that we can use to find the optimal decisions, that does not mean that there is no advice that the climate evidence can give to decision makers.

7.1. Uncertainty and decision making

We have seen that there are a variety of sources of uncertainty in climate modelling, and we have seen that the myriad ways of dealing with these uncertainties are often indirect and fraught with possible errors as well. There is no question on the basic science: increase in greenhouse gases will increase the temperature of the planet. This conclusion only requires some basic physics.[1] The uncertainties are all in what this conclusion means for local climate. Modern climate simulations are tremendously complex beasts: they have a horizontal spatial resolution of about 100 km which means the surface of the Earth is represented by about fifty thousand grid points. They have a vertical resolution varies much more, but thirty vertical levels is a reasonable headline figure. Temporal resolution is also quite variable, but it is of the order of about half an hour. The models track

[1] Recall the basic Energy-Balance Model we constructed in section 6.4.2.

something like a million variables, and involve about 100 adjustable parameters.[2] All of these components are interacting in myriad sophisticated ways (McGuffie and Henderson-Sellers 2005). What modern GCMs are not doing is working out whether climate change is a real phenomenon: It has been uncontroversial for a long while that increasing concentrations of greenhouse gases will cause an increase in global mean temperature. As far back as 1938 scientists were studying the effect CO_2 on temperature (Fleming 1998; Weart 2010).

If these conclusions are well-established, what are modern climate models *doing?* Knowing that annual global mean temperature will rise over the next fifty years doesn't tell us very much about what will happen to the climate on smaller scales: what will the weather be like in the Norwegian fjords, or in subsaharan Africa in this eventuality? How will temperature increase affect large scale atmospheric or oceanic phenomena like El Niño, the Gulf Stream or the hurricane season? How will rising temperatures affect number of rainy days in central London? Will climate change increase the risk of flooding in Oxford? These are the questions that climate models are now aiming to answer. Climate models are now part of the process of predicting finer-grained climate outcomes which are of more decision relevance to local and regional decision making bodies. The detailed climate evidence is used as input into more detailed models of regional climate and this in turn feeds in to assessments of the impacts of climate change. Each link in this chain of inferences has its own sources of uncertainties.

Let's say you are a policy maker who needs to decide some details of policy. The success of the policy will depend on the future evolution of the climate in some way. Let's say you are in charge of flood defences and you need to decide how high to build them. You can look at past data on maximum river height. This gives you some evidence about what appropriate defence heights should be. However, it would be naïve to stop there. Given the reality of global warming, you know that the climate state that will produce future floods is *relevantly different* from the state that produced the past data.

So you must turn now to projections of the future climate. Given the evidence of some corpus of climate models – say the crop of models discussed in the latest IPCC assessment report – what should you believe about the climate? As we have said, the basic physics is not in dispute and this alone sanctions general claims about increasing global mean temperature. We are interested in local effects. Let's

[2]These figures come from email conversations with Lenny Smith, Dave Stainforth and Erica Thompson.

say we are using something like the UK Climate Impacts Programme (UKCIP) data.

The UKCIP offers maps like Figure 7.1 for very fine grained predictions of precipitation seventy years into the future. A range scenarios (and their likelihoods) is offered (Jenkins et al. 2009, p. 17).

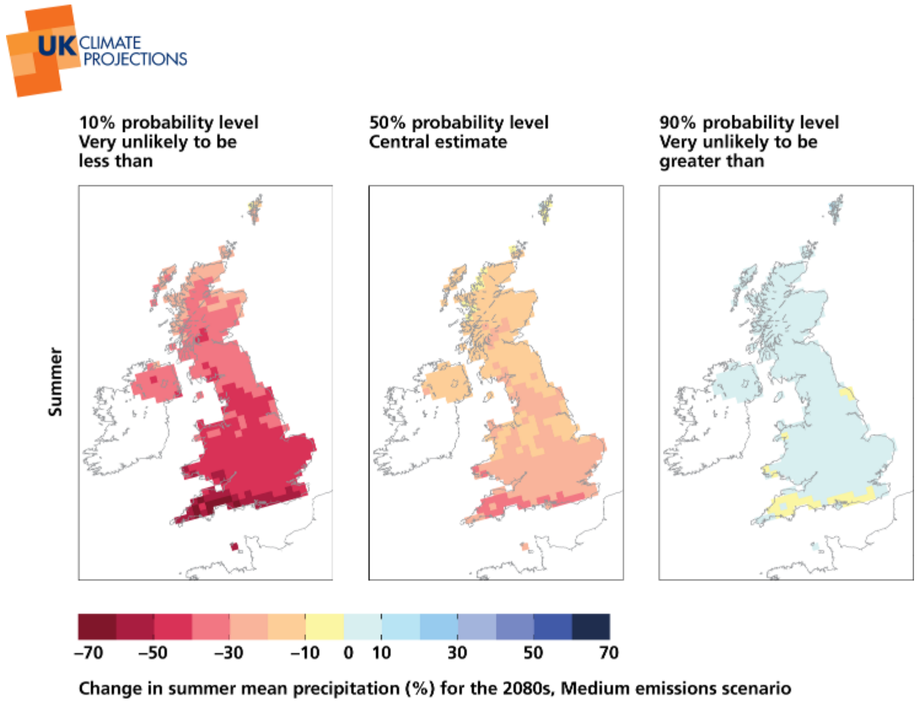

Figure 7.1.: UKCIP Projections for precipitation in 2080

For example, UKCP say this:

> The development of new techniques, together with increased computing power enabling them to be exploited, has allowed us to quantify the spread of future projections consistent with major known sources of uncertainty... Uncertainty in this case is dealt with by presenting projections which are probabilistic in nature. This sort of presentation is more informative than the single projection (for a given emissions scenario) in UKCIP02, or even a range of different projections from different climate models... but is also necessarily more complex. It gives the user the relative probability of different outcomes, based on the strength of evidence, and more openly reflects the state of the science. This is why probabilistic projections were adopted by IPCC for the first time in their most recent science assessment. The UKCP09 Projections respond to demands from a wide range of users for this

> level of detail. Jenkins et al. (2009, p. 13)

This probability then forms the basis of decisions that depend on the future evolution of the climate. That is, it is with respect to this ensemble probability that expected value is calculated, and decisions are taken that maximise this expectation. For example, perhaps the ensemble distribution is peaked around a small value for future precipitation in Oxford. On the basis of this prediction, flood insurance premiums in Oxfordshire are set relatively low, since the probability tells you that floods are unlikely. The worry about this sort of procedure is that the probabilistic projection doesn't account for all the possible sources of error discussed above. And this worry leads to scepticism in the use of these probabilities in decision support.

UKCIP's costings case studies, for example, use explicit probability numbers in their analysis:[3] The "Heritage building in Lewes" example depends on the following claim:

> By the 2080s, under the UKCIP02 medium-high emissions scenario, 7% of years will experience a wet "1994-5 type" winter.
> Taylor, Metroeconomica Ltd. And University of Bath (n.d. p. 7)

This probability forms the basis of a system of decision weights that is used to decide how much to spend adapting to these increased flood risks. So UKCIP really are selling these projections as giving you the probabilities of future climate outcomes.[4] Buried in UKCIP's "Risk, Uncertainty and Decision-making" handbook, there are some caveats and warnings, but they seem to get lost in what gets sold to policy makers. For instance

> The ability of mathematical risk models to handle a large number of complex interrelated issues is well tested. However, problems may be so large and complex that they cannot be resolved through the use of sophisticated models, although such models can still be of help in understanding the problem. Willows and Connell (2003, p. 28)

I claim that standard probability theory doesn't allow us to represent the kind of severe uncertainty we see in the case discussed here. The evidence we can glean from climate models is genuine evidence – so we are not in a state of ignorance

[3] `http://www.ukcip.org.uk/costings/case-studies/` retrieved 12 October 2011

[4] UKCIP don't make it clear how the ensemble output is dealt with in the process of producing the probabilistic projections. But as Frigg et al. (2013b) argue, probabilistic forecasts seem to have inherent problems.

– but it falls short of characterising an appropriate decision relevant probability function. So there is a question about what exactly we should believe on the basis of this evidence.

The last few chapters offered a formal theory of severe uncertainty and decision making. This theory is applicable in the climate case. This claim is backed up by, for example, the findings of Frigg et al. (2013b): probabilistic forecasters seem to be pragmatically worse off in a quite striking way.

7.2. More trouble for climate decisions

It gets worse before it gets better. As we have seen, climate science suffers from a number of difficult sources of error. But climate decisions also involve taking into account the difficulties of the economics of climate change. Weitzman (2009) points out some of these problems. First, given the long time that CO_2 stays in the atmosphere, it is very difficult to control the amount of it in the air. This *inertia* limits the options we have available. There is no reasonable method available to us that would even keep CO_2 levels constant.

Our evidence of the likelihood of the extreme events – the extreme costs we would like to avoid like catastrophic warming – are the events we have less evidence about. It is harder to learn about the tails of a distribution than it is to learn about those values around the mean that come up more often. It is the extremes of, for example, precipitation or temperature that will lead to the large costs that we want to avoid.

These are uncertainties at the global level. Then we have the unknown local responses to certain levels of global mean temperature change. These are the sorts of things that climate models are now trying to pin down.

Even if we had a good understanding of these facts about the future climate, there are uncertainties associated with how exactly these contingencies should be valued. What is the cost associated with this much extra flooding in Oxfordshire; with this sort of change in where storms make landfall on the Eastern seaboard of the United States? And that's just for the economic costs attached to certain kinds of possible climate contingencies. What about the unquantifiable human costs of hundreds of millions, possibly billions of people being displaced by climate induced extreme weather events and concomitant human problems; famines, wars and so on? I have in previous chapters just assumed that we can take precise utility functions for granted, but this is obviously a gross simplification. One problem with climate change and what the human response to it should be is that climate change happens on longer timescales than humans are used to dealing with. This typically means that the possible sacrifices we could make now will improve the lives of future generations. There are deep questions about how exactly we should value future generations.

Indeed, there are problems at an even deeper level. It isn't obvious that we have a good grasp of what acts are available to us in our climate decision problem. We don't know how effective the various policies on the table will be if implemented. We don't know how effective a particular carbon tax scheme will be at limiting emissions. We don't know how effective various technological solutions will be if funded. To put things in terms of a Savage-style decision problem, we don't know which functions from states to outcomes our policies match up with.

All of these things taken together prompt me to echo Weitzman's point that:

> such numbers and specifications *must* be imprecise and... this is a significant part of the climate-change economic-analysis problem, whose strong implications have thus far been ignored.
>
> Weitzman (2009, p. 9, Weitzman's emphasis)

And again:

> Perhaps in the end the climate-change economist can help most by not presenting a cost-benefit estimate for what is inherently [uncertain] as if it is accurate and objective... but instead by stressing somewhat more openly the fact that such an estimate might conceivably be arbitrarily *in*accurate depending upon what is subjectively assumed about [the loss function]. Even just acknowledging more openly the incredible magnitude of the deep structural uncertainties that are involved in climate-change analysis – and explaining better to policy makers that the artificial crispness conveyed by conventional IAM-based CBAs[5] here is especially and unusually misleading compared with more-ordinary non-climate-change CBA situations – might elevate the level of public discourse concerning what to do about global warming.
>
> Weitzman (2009, p. 26, Weitzman's emphasis)

Given these worries, we should be sceptical of strategies that focus too much on *optimising*: we don't know enough to properly determine what actions would

[5] Integrated Assessment Model-based Cost-Benefit Analyses.

be optimal. Imprecise probabilism refocuses attention on what we can do with respect to making choices: ruling out the bad acts. It may be that for certain kinds of climate decision, even this is too much to ask. There may be so much uncertainty surrounding a certain decision relevant question that only trivially worse acts are ruled out given the current state of evidence. I would argue that being stuck with such intractable decisions is better than optimising on the basis of too-sharp ensemble probability estimates. In the next section I discuss some kinds of strategy for dealing with even these apparently intractable decision problems. If the evidence really is that bad, then we are just stuck. What I would like to point out is that there could be cases where the evidence does not determine a decision relevant probability, but where it still provides enough guidance that some acts can be ruled out using some decision rule weaker than "optimise with respect to the decision relevant **pr**".

There are further problems with climate decisions, since there isn't just one agent in these scenarios. There is a strategic element to climate decisions: it doesn't matter if Britain commits to cutting CO_2 output by 20% if other countries don't agree to do the same. There is something of a "tragedy of the commons" about certain climate decisions. All countries want to emit CO_2 since industry, which generates wealth, emits CO_2. And let's say all countries know that we overall need to cut down on carbon output. But for each country, it would be better if the sacrifices came from *other* countries. Various international mechanisms for enforcing caps on emissions have proven to be fairly toothless, and unilateral action doesn't look appealing if the sacrifice isn't enough to mitigate the bad effects of climate change.

So climate decisions are difficult. I take one of the important things about the imprecise probabilities model I favour to be that it makes decision making difficult *when decision making should be difficult*, instead of seeming to make distinctions it shouldn't. Once it is emphasised that the cost-benefit analysis model seems to struggle to make appropriate discriminations, and thus fails to determine which acts are best given imprecise information, this leaves open the possibility for other kinds of concerns to help determine which acts will be the best ones. It is to these extra concerns that I turn in the next section.

One way we might go forward is to abandon the whole probabilistic framework and look at something different, like the "case-based decision theory" of Gilboa and Schmeidler (1995). The basic idea of this theory is instead of trying to construct a belief model of the current situation, one can solve a decision problem by looking at how successful your past actions have been in similar problems. Gilboa and Schmeidler offer a fairly complex mathematical theory of "problems" and the similarity relation between them. I have my doubts about the usefulness of this approach. First it relies on there being a history of similar problems that you or people you know have found themselves faced with. In the case of climate change, no such similar decisions have ever been faced. Never in the course of human history have we ever encountered a problem on this scale. One might be able to draw some analogies between climate change and smaller scale ecological policy, or with regulating nuclear power, or with phasing out CFCs. Despite certain similarities, none of these previous decisions is *similar* to the climate change problem in the way I think Gilboa and Schmeidler need. As an aside, I think that in cases where you do have similar cases, then you have enough previous information to get a moderately discriminating probabilistic model off the ground.

The imprecise probability model at least allows you to make some discriminations. It allows you to make choices *only* where the evidence sanctions the goodness of the choice. Consider deciding whether to offer tax breaks to green energy companies. There are many contingencies in play in this decision, but lets just focus on the possibility for green technology to slow global warming. The tax break is obviously a costly measure for the government[6] but the new green technology might stop severe warming and all the disastrous effects it would cause. The cost of the tax break is minuscule compared to the possible economic damage the unmitigated warming would cause. This looks like a case where the tax break option "almost dominates" the no tax break option, since the revenue from the extra taxes is basically irrelevant in the case of a global disaster.

This is not just a re-articulation of the famous *precautionary principle*, (Hansson 1997; Sandin et al. 2002; Steele 2006). I am not just saying that it is best to avoid actions (or omissions) that might cause massive harm. I am saying that some of the options available to us are so cheap compared to the harms that they might stop that they can be effected even based on the limited knowledge we now have. This sort of concern won't justify the large scale changes that have been advocated by, for example, the Stern Review (Stern 2007), but some beneficial actions can be justified in this way. But these large scale decisions aren't really my focus. On a global scale we know to some degree what will happen to the climate. I am

[6]Seen as an investment in a growth industry, it needn't even be viewed as a cost, but that is beside the point in the current context. Lord Stern has argued that the best way to make progress on climate change is to make green investments seem economically advantageous as well as ecologically.

interested more in the local-scale decisions like how high to build flood defences, how to modify building codes in areas at risk from tropical storms, how to set insurance premiums and so on. So what can we use to help with *those* sorts of decisions?

Before we move on, I should say something about whose utility function we are using in these decisions. Let's say you are deciding whether or not to implement a carbon tax. If you are CEO of a heavy industry company, then a carbon tax is going to be a significant cost, and the benefits of a carbon tax for yourself will be small. From an impersonal perspective, however, the economic damage of a carbon tax is offset by the huge benefits that will accrue to *future* generations who live in a significantly better world than the one they would have lived in had the tax not been implemented. So of course the value of an action depends on the utility function. The CEO's utility might be such that he prefers no tax, while the impersonal benevolent utilitarian's utility will likely favour the tax. In what follows, I am going to sweep all of the deep and interesting problems of the utility function under the carpet. My aim is to look at how *uncertainty and evidence* affect decision making. I think the important and difficult questions of what values we *should* be using when making these decisions – how to value future persons; how to spread the cost of mitigation – are orthogonal to the current project. One might want to use a similar "set of functions" approach to representing the multiplicity of values, and some of what I have said above will carry over straightforwardly. This is a worthwhile project, but not one I pursue here. Let's assume from now on that we are looking from the perspective of this impersonal benevolent utilitarian.

7.3. A more robust decision framework

We have seen how severe uncertainty both in the chances of climate contingencies, and in the value attached to those contingencies, makes orthodox decision making difficult or possibly inapplicable. We are faced then with a number of possible alternatives and no good way to choose between them. Let's say that we have at least ruled out the uncontroversially bad options with some sort of dominance reasoning. I ended chapter 5 with the claim that this is, essentially, as far as rationality can get you. This is an important point. Cost-benefit analysis based on spuriously precise probability estimates is not a better way to make decisions. I think this negative point is possibly more important than what little I have to offer by way of positive suggestions. The point is to emphasise that just because we can generate precise numbers doesn't mean that we can be confident of decisions made with them. And recognising this point – that decision making should be difficult – opens the way to an appropriate acknowledgement of the following sorts of concerns. My claim that severe uncertainty only allows for a weak standard of rationality – a standard that needn't determine which option to choose – is only half the story. There are cases where we need to make decisions, even though the evidence doesn't give us enough guidance. What criteria ought we adopt in these circumstances?

Stainforth et al. (2007b) offer a framework for using climate evidence to inform decision making. One of the important steps is working out whether the scientific evidence can usefully tell us about the variables of interest in the decision. At the level of the ensemble distribution, the ensemble is equally informative about all variables: it provides a marginal distribution for any variable it describes. So this step of deciding whether the evidence can tell us about the variables of relevance requires more than just looking at the ensemble distribution: it requires reflecting on which distributional features of the ensemble evidence can reasonably be taken to tell us something about the future evolution of the climate. Note the parallel with the earlier discussion of evidence and decision in the abstract: it is important that we allow our decision support mechanisms to fail to determine a best option when the evidence fails to support choosing one over the other. A naïve attitude to IAMs and ensemble evidence doesn't do this.

> The severity of model inadequacy suggests a more qualitative interpretation than one might wish. In particular, it is not at all clear that weighted combinations of results from today's complex climate models based on their ability to reproduce a set of observations can provide decision-relevant probabilities. Furthermore, they are liable to be misleading because the conclusions, usually in the form of PDFs, imply much greater confidence than the underlying assumptions justify.
>
> Stainforth et al. (2007a, p. 2158)

So what do we have left to help us decide what set of policies is best to pursue? I want to suggest a few values we might want to take into account: robustness, reversibility, adaptability. These are properties of acts that could be seen as *sui generis* good-making-features of those acts. This isn't to suggest that these are properties that cannot be folded into some sort of complicated expected utility calculation. They are features that are intuitively valuable, and recognisable as such outside of expected utility calculations. Given that we've seen that EU calculation becomes difficult if not impossible in cases of severe uncertainty, folding these features into an EU calculation would be a step backwards.

Lempert et al. (2004) offer an alternative framework which also seeks to build in a concern for the sorts of factors I discuss below. They contrast the standard *predict-then-act* approach with their favoured *assess-risk-of-policy* approach. The former is the standard procedure of optimising with respect to some probability. They characterise their alternative approach as follows:

> [T]he *assess-risk-of-policy* framework envisions a process that generates policy options whose satisfactory performance is maximally insensitive to uncertainties and outputs a small number of probabilities to characterize the residual risks of choosing such a policy.
>
> Lempert et al. (2004, p. 5)

They emphasise the value of insensitivity to uncertainties in the evidence: this relates to what I will call "robustness" below. They also emphasise the possible bad consequences of optimising with a possibly faulty probability.

How does the formal discussion of decision making in chapter 5 relate to the qualitative discussion of the current one? First both share a concern to represent uncertainty adequately; and this commitment entails allowing the decision making machinery to fail to deliver unambiguous answers. We might also see the concerns discussed below as being a particular instance of a set of tie-breaking heuristics that come into play in a particular context. Other sets of heuristics might come into play in other contexts.

7.3.1. Making decisions now

As more information comes in, as climate science improves, we will be able to make better decisions. Thus, we should make decisions now so as to facilitate the better decisions we will be able to make in the future. However, many decisions have to be made now. One thing we do know about the future of the climate is that the changes that will happen will be irreversible. Various inertias in the system (like how long CO_2 stays in the atmosphere) mean that it is already too late to stop some change, and the longer we wait to do the things we can to slow the change, the more change will be inevitable. The "wait and see" strategy of carrying on as normal until we have enough evidence to act decisively may lead to our being unable to change the course of the evolution of the climate. Better to cut emissions now, and then relax the constraints later as we learn more about what a reasonable "safe" level of emissions is; rather than leave emissions unchecked and run the risk of having to deal with GHGs with long residence times in the atmosphere.

The decisions we make now obviously can only rely on the evidence we currently have. We don't know enough to select some optimal strategy for possible future climates. What other criteria might we fall back on instead?

7.3.2. Robustness

There is some intuitive value to an act that leads to good outcomes a lot of the time. Take the example of building a bridge over a river. The changing climate might make the river more liable to flooding, or perhaps sea-level rise might cause the river level to be higher in the future. It seems like in this sort of circumstance, it is worthwhile to build bridges higher so that they would be above the level of the river even if the river level were to rise. This is related to the idea that acts that "nearly dominate" or are "ε away from dominating" *seem* like good acts.

Lots of possible actions – perhaps most actions – aren't sensitive to small changes in probability. In many cases, maximising with respect to some "middling" probability will produce an act that does *pretty well* in most cases. Consider the decision of whether to make building codes more stringent in an area that might be at increased risk of tropical storms due to climate change. In some scenarios, the area is at increased risk to storms, and the decision to tighten up building codes saves a great deal of money. In some scenarios, the storm risk stays the same or reduces, the costs associated with conforming to the more stringent building codes are smaller than the damages that would accrue in the previous scenarios. So as long as the probability of increased storm risk is *at least x*, the decision to tighten up the building regulations will be better than not doing so. This is a robust decision in the sense that the choice isn't sensitive to details of the probability distribution, as long as probability of increased risk is at least x. The value of x will depend on the values of the damages avoided and the costs accrued as discussed above. This is an example of a decision that can be taken without knowing the full PDF of possible futures. All that needs to be known is that the probability is at least *this much*. This is a criterion for decision making that relies on a weaker kind of evidence than optimising does. We don't need to know a full PDF, all we need is a lower bound on the probability. Recall the hierarchy of uncertainties from Kandlikar, Risbey, and Dessai (2005) that I mentioned in the introduction.

7.3.3. Reversibility

This virtue of possible strategies is often understood in the negative: we disvalue acts that are irreversible. One argument against many of the more extreme geo-engineering possibilities that have been mooted is that they are not reversible; or consider the debate on genetically modified (GM) crops. One argument against their use is that there's a danger that they could spread beyond the fields where they are being trialled. If we later learn of some bad side effect from having used these crops, we have no way of reining in their spread. This sort of irreversibility tends to be disvalued.[7]

It is a virtue of some act that if it turns out that it isn't effective, we can undo any adverse effects it might have. Let's say we learn that we are in a situation where some strategy has some bad side-effect which outweighs its benefits. We would prefer to return to the way things were before we effected this strategy: and therefore, to a situation where we don't have the bad side effect. Drastic measures will seem more palatable if we know that we can "rewind" things so that they are as they were before we started meddling. What reversibility of an action guarantees is that the "worst case scenario" for the action is only a little bit worse than the status quo.[8] Being only a little worse than the status quo in the worst case is a kind of "almost dominating" property.

7.3.4. Adaptability

Let's say you are in charge of flood defences. It seems a sensible strategy to build defences such that, if it turns out that larger defences are needed, the existing defences can easily be upgraded. This might involve making sure that people don't build too close to the defences, designing the foundations such that they can support bigger defences than are originally built on them, and so on.

This virtue, like the last, puts value on the possibility of "mid-course correction"; of being able to change tack in the light of new information. As our information improves, we hope to be able to make more targeted interventions. We should act now so as to allow our future decisions to be as effective as possible. Stainforth et al. (2007b) suggest this sort of thing when they say that even showing that the evidence doesn't support any particular option can be helpful:

[7] I am not endorsing this argument, merely using it as an example. As I understand it, GM crops are designed such that they cannot spread in the way this scenario imagines.

[8] It is the status quo, minus whatever costs are associated with reversing the effects of the previous action.

> Even this information can be useful to the decision maker in terms of focusing attention not on a particular response but on options which allow flexibility for adjustments in the future. (p.2174)

7.4. Conclusion

When evidence is incomplete, uncertain or error prone in the way that evidence from climate models is, we need an appropriately subtle and permissive representation of that uncertainty. This allows us to avoid spurious accuracy and spurious determinacy of choice. This leaves open the question of what choices should be made, among those options that cannot be ruled out on purely rational grounds. I have suggested a number of aspects of an option that might be considered valuable and that could feed into decision theory. This was done informally, because I think that the important conclusions of a previous chapter were that such decision making cannot be done purely formally.

Synopsis

I have argued that the multidimensional nature of uncertainty undermines the claim that standard probability theory is always the right formal model of belief and uncertainty. I have discussed several arguments for the claim that probability theory is the right model of belief and found flaws in each. Probabilism still serves as a regulative ideal, and the alternative framework I suggest is a conservative extension of it. I have discussed decision making with imprecise probabilities, although more work still needs to be done on this topic. There is no generally acceptable decisive imprecise decision rule: there are cases when rationality does not determine a choice. I concluded with a case study of uncertainty and decision making in climate science.

A. A proof of Joyce's theorem

A.1. Preliminary results

We are going to split the proof of the main theorem into several smaller parts. Joyce lists, without proof, six facts that allow him to prove his main theorem. I aim to provide the short proofs Joyce leaves out. First Joyce defines something like a notion of distance between belief functions.

Definition A.1.1 $\mathbf{D}(\mathbf{b},\mathbf{c}) = \mathbf{I}(\mathbf{v}+(\mathbf{b}-\mathbf{c}),\mathbf{v})$

I list these lemmas in order, so my Lemma A.1.1 corresponds to Joyce's Fact I and so on.

Lemma A.1.1 $\mathbf{D}(\bullet,\mathbf{c})$ *is continuous for each* $\mathbf{c} \in \mathbf{B}$

Proof If we fix some $\mathbf{c}$, it follows from Structure, that $\mathbf{D}(\mathbf{b},\mathbf{c}) = \mathbf{I}(\mathbf{v}+(\mathbf{b}-\mathbf{c}),\mathbf{v})$ is continuous for any $\mathbf{b}$. ∎

There are two places in the proof where continuity of $\mathbf{D}$ is important. One of them is just a use of the intermediate value theorem. For this, he doesn't really need continuity across the whole space, just continuity on lines and this follows from allowing the λ parameter to range over the reals. There is still the question of fundamentally relying on Euclidean distance, but that is unavoidable it seems. The other use of $\mathbf{D}$'s continuity is an application of the extreme value theorem. This relies on $\mathbf{V}^+$ being closed and bounded as well. This result needs more topological baggage.

Lemma A.1.2 $\mathbf{D}$*'s value does not depend on the choice of* $\mathbf{v} \in \mathbf{V}$.

Proof For all $X \in SL$ and $\mathbf{v},\mathbf{v}' \in \mathbf{V}$ with $\mathbf{b},\mathbf{c} \in \mathbf{B}$:

$$\begin{aligned} |\mathbf{v}(X) - (\mathbf{v}(X) + (\mathbf{b}(X) - \mathbf{c}(X)))| &= |\mathbf{b}(X) - \mathbf{c}(X)| \\ &= |\mathbf{v}'(X) - (\mathbf{v}'(X) + (\mathbf{b}(X) - \mathbf{c}(X)))| \end{aligned}$$

So, by Normality it follows that $\mathbf{I}(\mathbf{v}+\mathbf{b}-\mathbf{c},\mathbf{v}) = \mathbf{I}(\mathbf{v}'+\mathbf{b}-\mathbf{c},\mathbf{v}')$. So $\mathbf{D}$ does not depend on the choice of $\mathbf{v}$. ∎

Joyce has this fact following from Structure but Maher is right in pointing out that it actually follows from Normality .

Lemma A.1.3 $\mathbf{D}(\mathbf{b},\mathbf{c})$ *goes to infinity as* $\mathbf{b}(X)$ *goes to infinity for any* $X \in SL$.

Proof $\mathbf{I}(\mathbf{v}+(\mathbf{b}-\mathbf{c}),\mathbf{v})$ goes to infinity as $\mathbf{b}(X)$ goes to infinity for any X, by Structure. ∎

This is intuitive since $\mathbf{b}(X)$ tending to infinity is like moving away from the origin parallel to some axis. So the distance between **b** and some fixed **c** is going to increase in an unbounded way.

To prove the next result, we need something stronger than Dominance. Dominance says that if **b** and $\mathbf{b}'$ differ only at X, then which one is more accurate depends on which gets closer to the truth at X. What we need is: if **b** and $\mathbf{b}'$ differ only on some collection, $X_1, X_2, \dots X_n$, and one function is universally closer to the truth on those X_i, then *that* function is less inaccurate. What we really need is:

Strong Dominance *If* $\mathbf{b}(Y) = \mathbf{b}'(Y)$ *for all* $Y \notin \Phi \subseteq SL$ *and if* $|\mathbf{v}(X) - \mathbf{b}(X)| \geqslant |\mathbf{v}(X) - \mathbf{b}'(X)|$ *for all* $X \in \Phi$, *then* $\mathbf{I}(\mathbf{b},\mathbf{v}) \geqslant \mathbf{I}(\mathbf{b}',\mathbf{v})$.

This condition follows from Dominance by induction on the size of Φ. So we don't really need a stronger *axiom*, but it's worth pointing out this intermediate step in the proof.

Lemma A.1.4 $\mathbf{D}(\mathbf{b},\mathbf{c}) \geqslant \mathbf{D}(\mathbf{b}',\mathbf{c}')$ *if* $|\mathbf{b}(X) - \mathbf{c}(X)| \geqslant |\mathbf{b}'(X) - \mathbf{c}'(X)|$ *holds for all* $X \in SL$ *and the former equality is strict if the latter is strict for some* X.

Proof If $|\mathbf{b}(X) - \mathbf{c}(X)| \geqslant |\mathbf{b}'(X) - \mathbf{c}'(X)|$ then $|(\mathbf{v}(X) - \mathbf{v}(X)) + \mathbf{b}(X) - \mathbf{c}(X)| \geqslant |(\mathbf{v}(X) - \mathbf{v}(X)) + \mathbf{b}'(X) - \mathbf{c}'(X)|$. From this it follows by Dominance that $\mathbf{I}(\mathbf{v}+\mathbf{b}-\mathbf{c},\mathbf{v}) \geqslant \mathbf{I}(\mathbf{v}+\mathbf{b}'-\mathbf{c}',\mathbf{v})$. So $\mathbf{D}(\mathbf{b},\mathbf{c}) \geqslant \mathbf{D}(\mathbf{b}',\mathbf{c}')$. ∎

This lemma is saying that if **b** and **c** disagree more about everything than $\mathbf{b}'$ and $\mathbf{c}'$ do, then there is more distance between **b** and **c** than there is between $\mathbf{b}'$ and $\mathbf{c}'$.

Lemma A.1.5 *If* **c** *lies on* $\mathbf{b}\,\mathbf{b}'$, *and if* $\mathbf{c} \neq \mathbf{b}$ *then* $\mathbf{D}(\mathbf{b},\mathbf{b}') > \mathbf{D}(\mathbf{c},\mathbf{b}')$.

Proof $\mathbf{c} = \lambda\,\mathbf{b} + (1-\lambda)\,\mathbf{b}'$, for some $\lambda \in [0,1)$. For all X:

$$\begin{aligned}
|\mathbf{c}(X) - \mathbf{b}'(X)| &= |\lambda\,\mathbf{b}(X) + (1-\lambda)\,\mathbf{b}'(X) - \mathbf{b}'(X)| \\
&= |\lambda\,\mathbf{b}(X) - \lambda\,\mathbf{b}'(X)| \\
&= \lambda|\mathbf{b}(X) - \mathbf{b}'(X)| \\
&< |\mathbf{b}(X) - \mathbf{b}'(X)|
\end{aligned}$$

So, by Lemma A.1.4, we know $\mathbf{D}(\mathbf{b},\mathbf{b}') > \mathbf{D}(\mathbf{c},\mathbf{b}')$. ∎

Lemma A.1.6 $\mathbf{D}(\mathbf{b},\mathbf{c}) = \mathbf{D}(\mathbf{b}',\mathbf{c})$ *if and only if* $\mathbf{D}(\bullet,\mathbf{c})$ *has a unique minimum along* $\mathbf{b}\,\mathbf{b}'$ *at its midpoint* $\mathbf{m} = 1/2\,\mathbf{b} + 1/2\,\mathbf{b}'$.

Proof First, the "only if" part: $\mathbf{D}(\mathbf{b},\mathbf{c}) = \mathbf{D}(\mathbf{b}',\mathbf{c})$. So $\mathbf{I}(\mathbf{v}+\mathbf{b}-\mathbf{c},\mathbf{v}) = \mathbf{I}(\mathbf{v}+\mathbf{b}'-\mathbf{c},\mathbf{v})$. From Symmetry it follows that for any $\lambda \in [0,1]$:

$$\begin{aligned}&\mathbf{I}(\lambda[\mathbf{v}+\mathbf{b}-\mathbf{c}] + (1-\lambda)[\mathbf{v}+\mathbf{b}'-\mathbf{c}],\mathbf{v})\\ =\ &\mathbf{I}((1-\lambda)[\mathbf{v}+\mathbf{b}-\mathbf{c}] + \lambda[\mathbf{v}+\mathbf{b}'-\mathbf{c}],\mathbf{v})\end{aligned}$$

By Weak Convexity, it follows that this quantity is always greater than $\mathbf{I}(\mathbf{m},\mathbf{v})$ with equality only when $\lambda = 1/2$ (that is: the minimum is unique). ∎

As yet, I've found no proof of the "if" direction of this claim that Joyce makes. This direction is needed in the proof as Joyce sets it out, so this is a flaw in my presentation at the moment. That is, I currently have no proof that if $\mathbf{D}(\bullet,\mathbf{c})$ has a unique minimum on $\mathbf{b}\,\mathbf{b}'$ at $\mathbf{m}$ then $\mathbf{D}(\mathbf{b},\mathbf{c}) = \mathbf{D}(\mathbf{b}',\mathbf{c})$.

There are three more quick facts that Joyce uses, but does not state explicitly. For the sake of completeness, I offer them here.

Lemma A.1.7 $\mathbf{D}$ *is a symmetric function in its two arguments. That is,* $\mathbf{D}(\mathbf{b},\mathbf{c}) = \mathbf{D}(\mathbf{c},\mathbf{b})$ *for all* $\mathbf{b},\mathbf{c}$.

Proof $|\mathbf{b}(X) - \mathbf{c}(X)| = |\mathbf{c}(X) - \mathbf{b}(X)|$ for all X so by Lemma A.1.4 we have that $\mathbf{D}(\mathbf{b},\mathbf{c}) = \mathbf{D}(\mathbf{c},\mathbf{b})$. ∎

Lemma A.1.8 $\mathbf{I}(\mathbf{c},\mathbf{v}) = \mathbf{D}(\mathbf{c},\mathbf{v})$ *for all* $\mathbf{v} \in \mathbf{V}$ *and* $\mathbf{c} \in \mathbf{B}$.

Proof Follows from the definiton of $\mathbf{D}$ ∎

Lemma A.1.9 $\mathbf{D}(\mathbf{b},\mathbf{b})$ *is the* unique *minimum of* $\mathbf{D}(\bullet,\mathbf{b})$ *on any region containing* $\mathbf{b}$.

Proof $|\mathbf{b}(X) - \mathbf{b}(X)| = 0$ for all X. So for any $\mathbf{b}' \neq \mathbf{b}$, there is some X such that $\mathbf{b}'(X) \neq \mathbf{b}(X)$. From this it follows that $|\mathbf{b}(X) - \mathbf{b}(X)| < |\mathbf{b}(X) - \mathbf{b}'(X)|$ for that X. So by Lemma A.1.4, $\mathbf{D}(\mathbf{b},\mathbf{b}') > \mathbf{D}(\mathbf{b},\mathbf{b})$. ∎

A.2. Proving the "main theorem"

To reiterate, the theorem we are trying to prove is the following:

> (**Main Theorem**) If gradational inaccuracy is measured by a function $\mathbf{I}$ that satifies Structure, Extensionality, Dominance, Normality, Weak Convexity and Symmetry, then for each $\mathbf{c} \in \mathbf{B}\backslash\mathbf{V}^+$ there is a $\mathbf{c}' \in \mathbf{V}^+$ such that $\mathbf{I}(\mathbf{c},\mathbf{v}) > \mathbf{I}(\mathbf{c}',\mathbf{v})$ for *every* $\mathbf{v} \in \mathbf{V}$. (Joyce 1998, pp. 597–8)

So we are considering some agent with credences $\mathbf{c} \notin \mathbf{V}^+$. We break down the proof of the main theorem into three parts. The first part shows that there's a $\mathbf{c}'$ that uniquely minimises $\mathbf{D}(\bullet, \mathbf{c})$ on $\mathbf{V}^+$. The second part shows that there's an $\mathbf{m}$ such that for arbitrary $\mathbf{v}$, $\mathbf{I}(\mathbf{m}, \mathbf{v}) \geqslant \mathbf{I}(\mathbf{c}', \mathbf{v})$. And finally we show that $\mathbf{I}(\mathbf{c}, \mathbf{v}) > \mathbf{I}(\mathbf{m}, \mathbf{v})$. So, for every $\mathbf{c} \notin \mathbf{V}^+$, there's some $\mathbf{c}' \in \mathbf{V}^+$ that accuracy-dominates $\mathbf{c}$.

Theorem A.2.1 *There is a point $\mathbf{c}' \in \mathbf{V}^+$ such that the function $\mathbf{D}(\bullet, \mathbf{c})$ attains its* unique *minimum on $\mathbf{V}^+$ at $\mathbf{c}'$.*

Proof $\mathbf{V}^+$ is a closed and bounded set by definition of being a convex hull. $\mathbf{D}(\bullet, \mathbf{c})$ is a continuous (Lemma A.1.1) real-valued function (by definition). By a classic result from topology we know that a continuous real-valued function on a closed bounded region attains a minimum on that region. So $\mathbf{D}(\bullet, \mathbf{c})$ attains a minimum on $\mathbf{V}^+$. Call this minimum $\mathbf{c}'$.

We now need to show that $\mathbf{c}'$ is unique. Assume there were another minimum, $\mathbf{b}' \in \mathbf{V}^+$. $\mathbf{D}(\mathbf{c}', \mathbf{c}) = \mathbf{D}(\mathbf{b}', \mathbf{c})$. So by Lemma A.1.6, $\mathbf{D}(\bullet, \mathbf{c})$ attains a unique minimum on $\mathbf{c}'\mathbf{b}'$ at $1/2\,\mathbf{c}' + 1/2\,\mathbf{b}'$. $\mathbf{V}^+$ is convex and contains both $\mathbf{c}'$ and $\mathbf{b}'$, so it contains this mixture of them. That is $\mathbf{c}'\mathbf{b}' \subset \mathbf{V}^+$, so since $\mathbf{c}'$ is the unique minimum on $\mathbf{V}^+$ it must be the unique minimum on $\mathbf{c}'\mathbf{b}'$. So $\mathbf{c}'$ and $1/2\,\mathbf{c}' + 1/2\,\mathbf{b}'$ are both unique minima on $\mathbf{c}'\mathbf{b}'$. This is only possible if $\mathbf{b}' = \mathbf{c}'$. So $\mathbf{c}'$ must be a unique minimum. ∎

We have found our probability measure that will accuracy dominate $\mathbf{c}$. We need to now show that $\mathbf{c}'$ is indeed always more accurate than $\mathbf{c}$. There are two parts to this. First we construct an $\mathbf{m}$ which is less accurate than $\mathbf{c}'$. Then we show that this $\mathbf{m}$ is still more accurate than $\mathbf{c}$.

If $\mathbf{c}' = \mathbf{v}$, then it follows trivially from the strong version of Dominance that $\mathbf{I}(\mathbf{c}, \mathbf{v}) > \mathbf{I}(\mathbf{c}', \mathbf{v})$. Fix some arbitrary $\mathbf{v} \neq \mathbf{c}'$. We need to show the same inequality still holds.

Definition A.2.1 *Let $\mathbf{L} = \{\lambda\,\mathbf{c}' + (1-\lambda)\,\mathbf{v} : \lambda \in \mathbb{R}\}$. Let $\mathbf{L}_\lambda$ be the particular member of $\mathbf{L}$ with that value of λ. Let $\mathbf{R}$ be $\mathbf{L}$ restricted to $\lambda \geqslant 1$.*

$\mathbf{L}_0 = \mathbf{v}$ and $\mathbf{L}_1 = \mathbf{c}'$. $\mathbf{R}$ is the ray of $\mathbf{L}$ beginning at $\mathbf{c}'$ not containing $\mathbf{v}$. The idea of the proof is to find some $\mathbf{m}$ that we can compare easily with both $\mathbf{c}$ and $\mathbf{c}'$. We start by defining a line through $\mathbf{c}'$ and our arbitrary $\mathbf{v}$. Now, any $\mathbf{m}$ on the "$\mathbf{c}'$ side" of this line (that is, with $\lambda \geqslant 1$) will be less accurate than $\mathbf{c}'$. So all we need to do is find some such $\mathbf{m}$ that is easily comparable with $\mathbf{c}$ (See Figure A.1). This is probably best understood geometrically. Joyce can't help himself to all the ideas of geometry suggested by the pictures, but he can do some things. He can help himself to all of the "betweenness" notions of geometry on lines defined by mixtures (for example Hilbert's "Axioms of Order" Hilbert (1899)). The idea is that some point on the line between $\mathbf{v}$ and $\mathbf{k}$ must minimise the distance from $\mathbf{c}$ on that line. Geometrically, this is obvious since one can just "drop a perpendicular" from $\mathbf{c}$ onto the line.[1] Joyce can't really assume enough geometrical structure to make this obvious, so he has to proceed in a round about way.

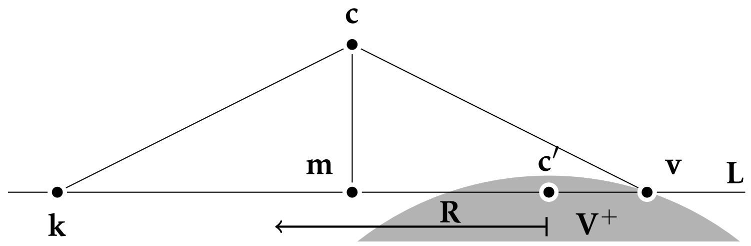

Figure A.1.: The geometry of Theorem A.2.2

Theorem A.2.2 *There is a point* $\mathbf{m} \in \mathbf{R}$ *such that (a)* $\mathbf{m}$ *uniquely minimises* $\mathbf{D}(\bullet,\mathbf{c})$ *on* $\mathbf{R}$, *(b)* $\mathbf{c}' \in \mathbf{m}\,\mathbf{v}$ *and (c)* $\mathbf{I}(\mathbf{m},\mathbf{v}) \geqslant \mathbf{I}(\mathbf{c}',\mathbf{v})$.

Proof There is an $X \in SL$ such that $\mathbf{v}(X) = 0$ and $\mathbf{c}'(X) \neq 0$ ($\mathbf{c}' \in \mathbf{V}^+$ so if $\mathbf{c}'$ and $\mathbf{v}$ agree on all zeroes, they agree on all ones too: they are identical). So $\mathbf{L}_\lambda(X) = \lambda\,\mathbf{c}'(X)$, which tends to infinity as λ does. By Lemma A.1.3 we have that $\mathbf{D}(\mathbf{L}_\lambda,\mathbf{c})$ goes to infinity as λ does.

$\mathbf{D}(\mathbf{c}',\mathbf{c}) < \mathbf{D}(\mathbf{v},\mathbf{c})$, since $\mathbf{c}'$ minimises $\mathbf{D}(\bullet,\mathbf{c})$ on $\mathbf{V}^+$. As a function of λ for $\lambda \geqslant 1$, $\mathbf{D}(\mathbf{L}_\lambda,\mathbf{c})$ ranges over at least the interval $[\mathbf{D}(\mathbf{c}',\mathbf{c}),\infty)$ and is continuous. So, by the intermediate value theorem, we know there is a λ such that $\mathbf{D}(\mathbf{L}_\lambda,\mathbf{c}) = \mathbf{D}(\mathbf{v},\mathbf{c})$. For this value of λ, let $\mathbf{L}_\lambda = \mathbf{k}$.

Let $\mathbf{m} = 1/2\,\mathbf{k} + 1/2\,\mathbf{v}$. By Lemma A.1.6 this $\mathbf{m}$ uniquely minimises $\mathbf{D}(\bullet,\mathbf{c})$ on $\mathbf{k}\,\mathbf{v}$. This proves (a).

$\mathbf{m}$ cannot be strictly[2] between $\mathbf{c}'$ and $\mathbf{v}$, because it would then be in $\mathbf{V}^+$, and this would contradict $\mathbf{c}'$ being the unique minimum of $\mathbf{D}(\bullet,\mathbf{c})$ in $\mathbf{V}^+$. That is, we know that $\mathbf{D}(\mathbf{m},\mathbf{c}) < \mathbf{D}(\mathbf{c}',\mathbf{c})$, since $\mathbf{c}' \in \mathbf{k}\,\mathbf{v}$. Note that all of $\mathbf{c}'\,\mathbf{v}$ is in $\mathbf{V}^+$. So if $\mathbf{m} \in \mathbf{c}'\,\mathbf{v}$, $\mathbf{m}$ would be in $\mathbf{V}^+$. But $\mathbf{c}'$ is the unique minimum of $\mathbf{D}(\bullet,\mathbf{c})$ in $\mathbf{V}^+$, so $\mathbf{m}$ can't be in there. So (b) $\mathbf{c}' \in \mathbf{m}\,\mathbf{v}$.

By Lemma A.1.5 and (b) we know that $\mathbf{D}(\mathbf{m},\mathbf{v}) > \mathbf{D}(\mathbf{c}',\mathbf{v})$. So $\mathbf{I}(\mathbf{v}+\mathbf{m}-\mathbf{v},\mathbf{v}) > \mathbf{I}(\mathbf{v}+\mathbf{c}'-\mathbf{v},\mathbf{v})$, and thus (c) $\mathbf{I}(\mathbf{m},\mathbf{v}) > \mathbf{I}(\mathbf{c}',\mathbf{v})$. ∎

[1] This way of speaking about the proof is due to Williams (2012b)

[2] Say $\mathbf{c}$ is strictly between $\mathbf{b}$ and $\mathbf{b}'$ if $\mathbf{c} \in \mathbf{b}\,\mathbf{b}'$ but $\mathbf{c} \neq \mathbf{b}$ and $\mathbf{c} \neq \mathbf{b}'$.

The next theorem can be understood by use of a similar diagram to before, but augmented to include some extra points. See Figure A.2.

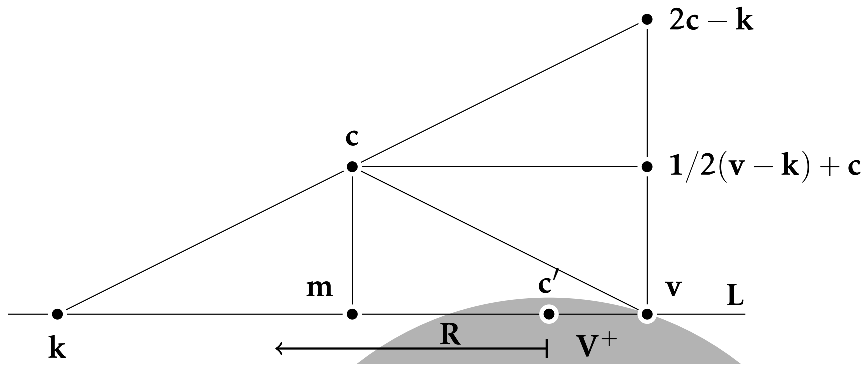

Figure A.2.: The geometry of Theorem A.2.3

Theorem A.2.3 $\mathbf{I}(\mathbf{c},\mathbf{v}) > \mathbf{I}(\mathbf{m},\mathbf{v})$

Proof By construction of $\mathbf{m}$ and $\mathbf{k}$ we know that $\mathbf{D}(\mathbf{k},\mathbf{c}) = \mathbf{D}(\mathbf{v},\mathbf{c})$. Now, consider the line from $\mathbf{k}$ to $2\mathbf{c}-\mathbf{k}$. The midpoint of this line is $\mathbf{c}$ and $\mathbf{D}(\bullet,\mathbf{c})$ attains a unique minimum at $\mathbf{c}$ (by Lemma A.1.9). So by the "if" direction of Lemma A.1.6, $\mathbf{D}(\mathbf{k},\mathbf{c}) = \mathbf{D}(2\mathbf{c}-\mathbf{k},\mathbf{c})$. These two equalities together give us: $\mathbf{D}(\mathbf{v},\mathbf{c}) = \mathbf{D}(2\mathbf{c}-\mathbf{k},\mathbf{c})$.

The other direction of Lemma A.1.6 (that is, the "only if" direction) then says that $\mathbf{D}(\bullet,\mathbf{c})$ attains a minimum at the midpoint of $\mathbf{v}$ and $2\mathbf{c}-\mathbf{k}$, namely at $1/2(\mathbf{v}-\mathbf{k})+\mathbf{c}$. By this result and Lemma A.1.7 we get that: $\mathbf{D}(\mathbf{c},\mathbf{v}) > \mathbf{D}(\mathbf{c},1/2(\mathbf{v}-\mathbf{k})+\mathbf{c})$.

$$
\begin{aligned}
\mathbf{I}(\mathbf{c},\mathbf{v}) &= \mathbf{D}(\mathbf{c},\mathbf{v}) && \text{By Lemma A.1.8}\\
&> \mathbf{D}(\mathbf{c},1/2(\mathbf{v}-\mathbf{k})+\mathbf{c}) && \text{by the above inequality}\\
&= \mathbf{I}(\mathbf{v}+[\mathbf{c}-1/2(\mathbf{v}-\mathbf{k})+\mathbf{c}],\mathbf{v}) && \text{By Definition A.1.1}\\
&= \mathbf{I}(1/2\,\mathbf{v}+1/2\,\mathbf{k},\mathbf{v}) &&\\
&= \mathbf{I}(\mathbf{m},\mathbf{v}) && \text{By definition of } \mathbf{m}
\end{aligned}
$$

So, $\mathbf{I}(\mathbf{c},\mathbf{v}) > \mathbf{I}(\mathbf{m},\mathbf{v})$. This concludes the proof. ∎

The three theorems together show that, if $\mathbf{c} \in \mathbf{B}\backslash\mathbf{V}^+$ then whatever $\mathbf{I}$ you use, there is some $\mathbf{c}' \in \mathbf{V}^+$ such that, for all $\mathbf{v} \in \mathbf{V}$, $\mathbf{I}(\mathbf{c},\mathbf{v}) > \mathbf{I}(\mathbf{c}',\mathbf{v})$.

Bibliography

The following figures are from Solomon et al. (2007). Figure 6.1 is FAQ1.2 Figure 1; Figure 6.3 is Figure 1.2; Figure 6.5 is part of Figure 1.4; Figure 6.6 is FAQ 1.1 Figure 1. Figures downloaded from `http://www.ipcc.ch/`. Figure 7.1 was downloaded from `http://www.ukcip.org.uk/wordpress/wp-content/UKCP09/Summ_Pmean_med_2080s.png` on 12 October 2011, it is © UKCP 2009. All other figures drawn with R or TikZ by the author. The trick dice are from the CATS/Grantham Institute stand at a Royal Society event. Figures A.1 and A.2 after Joyce (1998).

Akaike, H. (1973). "Information Theory as an Extension of the Maximum Likelihood Principle". In: *Second International Symposium on Information Theory*. Ed. by B. Petrov and F. Csaki. Akademiai Kiado, pp. 267–281.

Allingham, M. (2002). *Choice Theory: A very short introduction*. Oxford Unversity Press.

Anscombe, F. J. and R. J. Aumann (1963). "A Definition of Subjective Probability". *The Annals of Mathematical Statistics* 34, pp. 199–205.

Arnborg, S. and G. Sjödin (2000). "Bayes Rules in Finite Models". In: *Proceedings of the European Conference on Artificial Intelligence*, pp. 571–575.

Bandyopadhyay, P. S. (1994). "In Search of a Pointless Decision Principle". *Philosophy of Science Association Proceedings*, pp. 260–269.

Bermúdez, J. L. (2009). *Decision Theory and Rationality*. Oxford University Press.

Binmore, K. (2008). *Rational Decisions*. Princeton University Press.

Borges, J. L. (1999). "On Exactitude in Science". In: *Borges Collected Fictions*. Translated by Andrew Hurley. Penguin, p. 325.

Bovens, L. and S. Hartmann (2003). *Bayesian epistemology*. Oxford University Press.

Bradley, D. and H. Leitgeb (2006). "When betting odds and credences come apart: More worries for Dutch book arguments". *Analysis* 66, pp. 119–127.

Bradley, R. (1998). "A representation theorem for decision theory with conditionals". *Synthese* 116, pp. 187–229.

——— (2004). "Ramsey's Representation Theorem". *Dialectica* 58, pp. 483–498.

Bradley, R. (2007). “A Unified Bayesian Decision Theory”. *Theory and Decision* 63, pp. 233–263.

——— (2009). “Revising Incomplete Attitudes”. *Synthese* 171, pp. 235–256.

Bradley, S. (2011). *Scientific Uncertainty: A User’s Guide*. Tech. rep. 56. Grantham Institute Working Paper.

——— (2012). “Dutch book arguments and imprecise probabilities”. In: *Probabilities, Laws and Structures*. Ed. by D. Dieks, W. J. González, S. Hartmann, M. Stöltzner, and M. Weber. Springer, pp. 3–17.

Bradley, S. and K. Steele (ms.[a]). *Can Free Evidence be Bad? Value of Information for the Imprecise Probabilist*.

——— (ms.[b]). *Subjective Probabilities Need Not Be Sharp*.

——— (ms.[c]). *Uncertainty, Learning and the "Problem" of Dilation*.

Bröcker, J. (2009). “Reliability, Sufficiency and the Decomposition of Proper Scores”. *The Quarterly Journal of the Royal Meteorological Society* 135, pp. 1512–1519.

Bröcker, J. and L. A. Smith (2007). “Scoring Probabilistic Forecasts; On the Importance of Being Proper”. *Weather and Forecasting* 22, pp. 382–388.

Broome, J. (1991). *Weighing Goods*. Blackwell.

Buchak, L. (ms.). *Risk aversion and rationality*. manuscript.

Camerer, C. and M. Weber (1992). “Recent Developments in Modeling Preferences: Uncertainty and Ambiguity”. *Journal of Risk and Uncertainty* 5, pp. 325–370.

Chang, H. (2004). *Inventing Temperature*. Oxford University Press.

Christensen, D. (2001). “Preference-based arguments for probabilism”. *Philosophy of Science* 68, pp. 356–376.

——— (2004). *Putting logic in its place*. Oxford University Press.

——— (2009). “Disagreement as Evidence: The Epistemology of Controversy”. *Philosophy Compass*, pp. 756–767.

Chu, F. and J. Halpern (2004). “Great expectations. Part II: Generalized expected utility as a universal decision rule”. *Artificial intelligence* 159, pp. 207–230.

——— (2008). “Great expectations. Part I: On the customizability of General Expected Utility”. *Theory and Decision* 64, pp. 1–36.

Clifford, W. K. (1901). “The Ethics of Belief”. In: *Lectures and Essays*. Ed. by L. Stephen and F. Pollock. Third Edition. Vol. 2. Macmillan, pp. 161–205.

Colyvan, M. (2008). “Relative Expectation Theory”. *The Journal of Philosophy* 105, pp. 37–44.

Covey, C. (2000). "Beware the Elegance of the Number Zero". *Climactic Change* 44, pp. 409–411.

Cox, R. T. (1946). "Probability, Frequency and Reasonable Expectation". *American Journal of Physics* 14, pp. 1–13.

Cozman, F. (2012). "Sets of probability distributions, independence and convexity". *Synthese* 186, pp. 577–600.

——— (n.d.). *A Brief Introduction to the Theory of Sets of Probability Measures*. http://www.poli.usp.br/p/fabio.cozman/Research/CredalSetsTutorial/quasi-bayesian.html.

Dodd, D. (forthcoming). "Roger White's Argument Against Imprecise Credences". *British Journal for the Philosophy of Science*.

Döring, F. (2000). "Conditional Probability and Dutch Books". *Philosophy of Science* 67, pp. 391–409.

Duhem, P. (1998). "Physical Theory and Experiment". In: *Philosophy of Science, The Central Issues*. Ed. by M. Curd and J. Cover. W.W. Norton & Co., pp. 257–279.

Earman, J. (1986). *A primer on determinism*. D. Reidel Publishing Company.

Eichenberger, J. and D. Kelsey (2009). "Ambiguity". In: *The Handbook of Rational and Social choice*. Ed. by P. Anand, P. Pattanaik, and C. Puppe. Oxford University Press. Chap. 4, pp. 113–139.

Elga, A. (2000). "Self-locating Belief and the Sleeping Beauty Problem". *Analysis* 60, pp. 143–147.

——— (2010). "Subjective Probabilities should be Sharp". *Philosophers' Imprint* 10.

Ellis, B. D. (1966). *Basic concepts of measurement*. Cambridge University Press.

Eriksson, L. and A. Hájek (2007). "What Are Degrees of Belief?" *Studia Logica* 86, pp. 183–213.

Fefferman, C. L. (n.d.). *The Navier-Stokes equation*. http://www.claymath.org/millennium/Navier-Stokes_Equations/navierstokes.pdf.

Fine, T. (1973). *Theories of Probability: An Examination of Foundations*. Academic Press.

Fishburn, P. (1973). "A Mixture-Set Axiomatization of Conditional Subjective Expected Utility". *Econometrica* 41, pp. 1–25.

Fleming, J. R. (1998). *Historical Perspectives on Climate Changes*. Oxford University Press.

Frigg, R. (2008a). "A field guide to recent work on the foundations of statistical mechanics". In: *The Ashgate companion to contemporary philosophy of physics*. Ed. by D. Rickles. Ashgate, pp. 99–196.

——— (2008b). "Humean chance in Boltzmannian statistical mechanics". *Philosophy of Science* 75, pp. 670–681.

Frigg, R. and J. Reiss (2009). "Philosophy of Simulation: Hot New Issues or Same Old Stew?" *Synthese* 169, pp. 593–613.

Frigg, R., S. Bradley, H. Du, and L. A. Smith (2013a). *Laplace's Demon and Climate Change*. Tech. rep. 103. Grantham Institute Working Paper.

Frigg, R., S. Bradley, R. L. Machete, and L. A. Smith (2013b). "Probabilistic Forecasting: Why Model Imperfection is a Poison Pill". In: *New Challenges to Philosophy of Science*. Ed. by H. Andersen, D. Dieks, G. Wheeler, W. Gonzalez, and T. Uebel. Springer.

Gaertner, W. (2009). *A Primer in Social Choice Theory*. Oxford University Press.

Gärdenfors, P. and N.-E. Sahlin (1982). "Unreliable probabilities, risk taking and decision making". *Synthese* 53, pp. 361–386.

Gibbard, A. (2007). "Rational credence and the value of truth". *Oxford Studies in Epistemology*, pp. 143–164.

Gilboa, I. and D. Schmeidler (1993). "Updating ambiguous beliefs". *Journal of Economic Theory* 59, pp. 33–49.

——— (1995). "Case-Based Decision Theory". *The Quarterly Journal of Economics* 110, pp. 605–639.

Grove, A. and J. Y. Halpern (1998). "Updating sets of probabilities". In: *Proceedings of the fourteenth conference on uncertainty in AI*, pp. 173–182.

Haenni, R., J.-W. Romeijn, G. Wheeler, and J. Williamson (2010). *Probabilistic Logic and Probabilistic Networks*. Synthese Library.

Hájek, A. (2003). "What conditional probabilities could not be". *Synthese* 137, pp. 273–323.

——— (2007). "The reference class problem is your problem too". *Synthese* 156, pp. 563–585.

——— (2008). "Arguments for—or against—probabilism?" *British Journal for the Philosophy of Science* 59, pp. 793–819.

——— (ms.). *Staying Regular?*

Hájek, A. and M. Smithson (2012). "Rationality and Indeterminate Probabilities". *Synthese* 187, pp. 33–48.

Halpern, J. Y. (1999a). "A counterexample to theorems of Cox and Fine". *Journal of Artificial Intelligence Research* 10, pp. 67–85.

——— (1999b). "Cox's theorem revisited". *Journal of Artificial Intelligence Research* 11, pp. 429–435.

——— (2003). *Reasoning about uncertainty*. MIT press.

Hansson, S. O. (1997). "The Limits of Precaution". *Foundations of Science* 2, pp. 293–306.

——— (2011). "Logic of Belief Revision". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta.

Hawthorne, J. (2005). "Degree-of-Belief and Degree-of-Support: Why Bayesians Need Both Notions". *Mind* 114, pp. 277–320.

——— (2009). "The Lockean Thesis and the Logic of Belief". In: *Degrees of Belief*. Ed. by F. Huber and C. Schmidt-Petri. Springer, pp. 49–74.

Herron, T., T. Seidenfeld, and L. Wasserman (1994). "The Extent of Dilation of Sets of Probabilities and the Asymptotics of Robust Bayesian Inference". In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp. 250–259.

Hilbert, D. (1899). *The foundations of geometry*. Available from Project Gutenberg `http://www.gutenberg.org/etext/17384`. Open Court Publishing.

Hitchcock, C. and E. Sober (2004). "Prediction Versus Accommodation and the Risk of Overfitting". *British Journal for the Philosophy of Science*, pp. 1–34.

Hodges, J. and E. Lehman (1951). "The Uses of Previous Experience in Reaching Statistical Decisions". *Annals of Mathematical Statistics* 23, pp. 396–407.

Hosni, H. (ms.). *Towards a Bayesian Theory of Second Order Uncertainty*.

Houghton, J. (2009). *Global Warming: The Complete Briefing*. Fourth. Cambridge University Press.

Howson, C. (2009). "Epistemic Probability and Coherent Degrees of Belief". In: *Degrees of Belief*. Ed. by F. Huber and C. Schmidt-Petri. Springer, pp. 97–119.

——— (2012). "Modelling Uncertain Inference". *Synthese* 186, pp. 475–492.

Hsieh, N.-h. (2008). "Incommensurable Values". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2008.

Huber, F. (2009). "Belief and Degrees of Belief". In: *Degrees of Belief*. Ed. by F. Huber and C. Schmidt-Petri. Springer, pp. 1–33.

Huber, F. and C. Schmidt-Petri, eds. (2009). *Degrees of Belief*. Springer.

Hume, D. (1975 [1777]). *Enquiries concerning human understanding*. Third Edition. Oxford University Press.

Humphreys, P. W. (1985). "Why propensities cannot be probabilities". *The Philosophical Review* 94, pp. 557–570.

Jaffray, J.-Y. (1989). "Coherent bets under partially resolving uncertainty and belief functions". *Theory and Decision* 26, pp. 90–105.

James, W. (1897). "The Will to Believe". In: *The Will to Believe and other essays in popular philosophy*. Longmans, Green and Co., pp. 1–31.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Early draft version available: http://omega.math.albany.edu:8008/JaynesBook.html. Cambridge University Press.

Jeffrey, R. (1983). *The logic of decision*. 2nd ed. University of Chicago Press.

Jenkins, G., J. Murphy, D. Sexton, J. Lowe, P. Jones, and C. Kilsby (2009). *UKCP Briefing Report*. http://ukclimateprojections.defra.gov.uk/images/stories/briefing_pdfs/UKCP09_Briefing.pdf.

Joyce, J. M. (1998). "A Nonpragmatic Vindication of Probabilism". *Philosophy of Science* 65, pp. 575–603.

——— (1999). *The foundations of causal decision theory*. Cambridge studies in probability, induction and decision theory. Cambridge University Press.

——— (2005). "How Probabilities Reflect Evidence". *Philosophical Perspectives* 19, pp. 153–178.

——— (2009). "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief". In: *Degrees of Belief*. Ed. by F. Huber and C. Schmidt-Petri. Springer, pp. 263–297.

——— (2011). "A Defense of Imprecise Credence". *Oxford Studies in Epistemology* 4. forthcoming.

Kandlikar, M., J. Risbey, and S. Dessai (2005). "Representing and communicating deep uncertainty in climate-change assessments". *Comptes Rendus Geoscience* 337, pp. 443–455.

Kaplan, M. (1996). *Decision Theory as Philosophy*. Cambridge University Press.

——— (2010). "In defense of modest probabilism". *Synthese* 176, pp. 41–55.

Keynes, J. M. (1921). *A treatise on probability*. Macmillan.

Keynes, J. M. (1923). *A Tract on Monetary Reform*. MacMillan.

Knight, F. (1921). *Risk, Uncertainty and Profit*. Houghton and Mifflin.

Koopman, B. O. (1940). "The Bases of Probability". *Bulletin of the American Mathematical Society* 46, pp. 763–774.

Krantz, D, R. D. Luce, A. Tversky, and P. Suppes (1971). *Foundations of Measurement Volume I: Additive and Polynomial Representations*. Dover Publications.

Kreps, D. M. (1988). *Notes on the Theory of Choice*. Westview Press.

Kyburg, H. E. (1983). "Rational belief". *The Brain and Behavioural Sciences* 6, pp. 231–273.

——— (1984). *Theory and Measurement*. Cambridge University Press.

Kyburg, H. E. and M. Pittarelli (1992). *Set-based Bayesianism*. Tech. rep. UR CSD;TR407. `http://hdl.handle.net/1802/765`. University of Rochester, Computer Science Department.

Kyburg, H. E. and C. M. Teng (2001). *Uncertain Inference*. Cambridge University Press.

Lakatos, I. (1976). *Proofs and Refutations*. Cambridge University Press.

Laplace, P. S. d. (1951 [1816]). *A Philosophical Essay on Probabilities*. Translated from the French by Frederick Wilson Truscott and Frederick Lincoln Emery. Dover.

Leitgeb, H. and R. Pettigrew (2010a). "An Objective Justification of Bayesianism I: Measuring Inaccuracy". *Philosophy of Science* 77, pp. 201–235.

——— (2010b). "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy". *Philosophy of Science* 77, pp. 236–272.

Lempert, R., N. Nakicenovic, D. Sarewitz, and M. Schlesinger (2004). "Characterizing Climate-Change Uncertainties for Decision-Makers". *Climactic Change* 65, pp. 1–9.

Levi, I. (1974). "On Indeterminate probabilities". *Journal of Philosophy* 71, pp. 391–418.

——— (1986). *Hard choices: decision making under unresolved conflict*. Cambridge University Press.

Lewis, D. (1986). "A Subjectivist's Guide to Objective Chance (and postscript)". In: *Philosophical Papers II*. Oxford University Press, pp. 83–132.

——— (1994). "Humean Supervenience Debugged". *Mind* 103, pp. 473–490.

Lloyd, E. A. (2009). "Varieties of Support and Confirmation of Climate Models". *Proceedings of the Aristotelian Society Supplementary Volume* LXXXIII, pp. 213–232.

Lo, A. W. and M. T. Mueller (2010). *Warning: Physics Envy may be Hazardous to your Wealth*. `arXiv:1003.2688v3`.

Loewer, B. (2001). "Determinism and chance". *Studies in the History and Philosophy of Modern Physics* 32, pp. 609–620.

——— (2004). "David Lewis's Humean Theory of Objective Chance". *Philosophy of Science* 71, pp. 1115–1125.

Luce, R. D. and H. Raiffa (1989). *Games and Decisions*. Dover.

Maher, P. (1993). *Betting on Theories*. Cambridge University Press.

——— (2002). "Joyce's Argument for Probabilism". *Philosophy of Science* 69, pp. 73–81.

May, R. (1976). "Simple mathematical models with very complex dynamics". *Nature* 261, pp. 459–467.

McGuffie, K. and A. Henderson-Sellers (2005). *A Climate Modelling Primer*. Third Edition. Wiley.

Meacham, C. and J. Weisberg (2011). "Representation Theorems and the Foundations of Decision Theory". *Australasian Journal of Philosophy* 89, pp. 641–663.

Milnor, J. (1951). *Games against Nature*. Tech. rep. RAND corporation.

Muldoon, R. (2007). "Robust Simulation". *Philosophy of Science* 74, pp. 873–883.

Murphy, J., B. Booth, M. Collins, G. Harris, D. Sexton, and M. Webb (2007). "A methodology for proabilistic predictions of regional climate change from perturbed physics ensembles". *Philosophical Transactions of the Royal Society A* 365, pp. 1993–2028.

Norton, J. (2003). "A Material Theory of Induction". *Philosophy of Science* 70, pp. 647–670.

Okasha, S. (2007). "Rational choice, risk aversion and rationality". *Journal of Philosophy* 104, pp. 217–235.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994). "Verification, Validation and Confirmation of Numerical Models in the Earth Sciences". *Science* 263, pp. 641–646.

Paris, J. (1994). *The uncertain reasoner's companion*. Cambridge University Press.

——— (2005 [2001]). "A note on the Dutch book method". In: *Proceedings of the Second International Symposium on Imprecise Probabilities and their Applications*, pp. 301–306.

Parker, W. (2006). "Understanding Pluralism in Climate Modeling". *Foundations of Science* 11, pp. 349–368.

——— (2009). "Confirmation and Adequacy-for-Purpose in Climate Modeling". *Proceedings of the Aristotelian Society Supplementary Volume* LXXXIII, pp. 233–249.

——— (2010a). "Predicting weather and climate: Uncertainty, ensembles and probability". *Studies in History and Philosophy of Modern Physics* 41, pp. 263–272.

Parker, W. (2010b). "Whose Probabilities? Predicting Climate Change with Ensembles of Models". *Philosophy of Science* 77, pp. 985–997.

Petersen, A. (2000). "Philosophy of Climate Science". *Bulletin of the American Meteorological Society*, pp. 265–271.

——— (2012). *Simulating Nature: A Philosophical Study of Computer Simulation and Their Role in Climate Science and Policy Advice*. Second. CRC Press/Taylor & Francis.

Peterson, M. (2009). *An Introduction to Decision Theory*. Cambridge University Press.

Pettigrew, R. (2011). "Epistemic Utility Arguments for Probabilism". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta.

——— (2012). "Accuracy, chance and the Principal Principle". *Philosophical Review* 121, pp. 241–275.

Petzold, C. (1999). *Code: The hidden language of computer hardware and software*. Microsoft Press.

Pierrehumbert, R. T. (2011). *Principles of Planetary Climate*. Chapter 3 available online here: `http://www-das.uwyo.edu/~deshler/Atsc4400_5400_Climate/PierreHumbert_Climate_Ch3.pdf`. Cambridge University Press.

Popper, K. (1982). *The Open Universe*. Routledge.

Predd, J., R. Seiringer, E. H. Lieb, D. Osherson, H. V. Poor, and S. Kulkarni (2009). "Probabilistic Coherence and proper scoring rules". *IEEE Transactions on information theory* 55, pp. 4786–4792.

Ramsey, F. P. (1926). "Truth and Probability". In: *The Foundations of Mathematics and other Logical Essays*. 1999 electronic edition. Routledge, pp. 156–198.

Rédei, M. (1998). *Quantum Logic in Algebraic Approach*. Kluwer Academic Publishers.

Rédei, M. and S. Summers (2007). "Quantum probability theory". *Studies in the History and Philosophy of Modern Physics* 38, pp. 390–417.

Regan, H., M. Colyvan, and M. A. Burgman (2002). "A Taxonomy and Treatment of Uncertainty for Ecology and Conservation Biology". *Ecological Applications* 12, pp. 618–628.

Sandin, P., M. Peterson, S. O. Hansson, C. Rudén, and A. Juthe (2002). "Five charges against the precautionary principle". *Journal of Risk Research* 5, pp. 287–299.

Savage, L. (1972 [1954]). *The Foundations of Statistics*. 2nd ed. Dover.

Scaife, A. A., T. Spangehl, D. R. Fereday, U. Cubasch, U. Langematz, H. Akiyoshi, S. Bekki, P. Baesicke, N. Butchart, M. P. Chipperfiled, A. Gettelman, S. C. Hardiman, M. Michou, E. Rozanov, and T. G. Shepherd (2012). "Climate Change Projections and Statosphere-Troposphere Interaction". *Climate Dynamics* 38, pp. 2089–2097.

Schaffer, J. (2007). "Deterministic Chance?" *British Journal for the Philosophy of Science* 58, pp. 114–140.

Schervish, M., T. Seidenfeld, and J. B. Kadane (1990). "State-Dependent Utilities". *Journal of the American Statistical Association* 85, pp. 840–847.

Schick, F. (1986). "Dutch Bookies and Money Pumps". *Journal of Philosophy* 83, pp. 112–119.

Schiermeier, Q. (2010). "The Real Holes in Climate Science". *Nature* 463, pp. 284–287.

Schwarz, W. (ms.). *Lost Memories and Useless Coins: Revisiting the Absentminded Driver*. `https://github.com/wo/papers/blob/master/driver.pdf?raw=true`.

Seidenfeld, T. (1993). "Outline of a Theory of Partially Ordered Preferences". *Philosophical Topics* 21, pp. 173–188.

——— (2004). "A contrast between two decision rules for use with (convex) sets of probabilities: Γ-maximin versus *E*-admissibility". *Synthese* 140, pp. 69–88.

Seidenfeld, T., M. Schervish, and J. Kadane (1995). "A Representation of Partially Ordered Preferences". *Annals of Statistics* 23, pp. 2168–2217.

Seidenfeld, T. and L. Wasserman (1993). "Dilation for sets of probabilities". *Annals of Statistics* 21, pp. 1139–1154.

Sen, A. (1970). *Collective choice and social welfare*. Holden-Day.

——— (1977). "Social choice theory: A re-examination". *Econometrica* 45, pp. 53–89.

Skyrms, B. (2011). "Resiliency, Propensities and Causal Necessity". In: *Philosophy of Probability: Contemporary Readings*. Ed. by A. Eagle. Routledge, pp. 529–536.

Smets, P. and R. Kennes (1994). "The Transferable Belief Model". *Artificial Intelligence* 66, pp. 191–234.

Smith, L. (1996). "Accountability in ensemble prediction". In: *Proceedings of the ECMWF workshop on predictability*, pp. 351–368.

——— (2000). "Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems". In: *Nonlinear Dynamics and Statistics*. Ed. by A. Mees. Birkhauser.

——— (2007). *Chaos: A very short introduction*. Oxford University Press.

Smith, P. (1998). *Explaining chaos*. Cambridge University Press.

Snow, P. (1998). "On the Correctness and Reasonableness of Cox's Theorem for Finite Domains". *Computational Intelligence* 14, pp. 452–459.

Sober, E. (2002). "Instrumentalism, Parsimony and the Akaike Framework". *Philosophy of Science* 69, S112–S123.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Trignor, and H. Miller, eds. (2007). *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Sorensen, R. (2012). "Vagueness". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta.

Spolsky, J. (2002). *The law of leaky abstractions*. `http://www.joelonsoftware.com/articles/LeakyAbstractions.html`.

Stainforth, D. A., M. R. Allen, E. Tredger, and L. A. Smith (2007a). "Confidence uncertainty and decision-support relevance in climate models". *Philosophical Transactions of the Royal Society* 365, pp. 2145–2161.

Stainforth, D. A., T. Downing, R. Washington, A. Lopez, and M. New (2007b). "Issues in the interpretation of climate model ensembles to inform decisions". *Philosophical Transactions of the Royal Society* 365, pp. 2163–2177.

Steele, K. (2006). "The precautionary principle: A new approach to public decision-making?" *Law, Probability and Risk* 5, pp. 19–31.

——— (2007). "Distinguishing indeterminate belief from "risk averse" preference". *Synthese* 158, pp. 189–205.

——— (2010). "What are the minimal requirements of rational choice? Arguments from the sequential setting". *Theory and Decision* 68.4, pp. 463–487.

Stegenga, J. (2009). "Robustness, Discordance and Relevance". *Philosophy of Science* 76, pp. 650–661.

Stern, N. (2007). *The Economics of Climate Change*. Cambridge University Press.

Stevens, S. S. (1946). "On the theory of Scales of Measurement". *Science* 103, pp. 677–680.

Stirling, A. (2010). "Keep it complex". *Nature* 468, pp. 1029–1031.

Sturgeon, S. (2008). "Reason and the grain of belief". *Noûs* 42, pp. 139–165.

Suárez, M. (2007). "Quantum Propensities". *Studies in the History and Philosophy of Modern Physics* 38, pp. 418–438.

——— (forthcoming). "Propensities and Pragmatism". *Journal of Philosophy*.

Sugden, R. (1985). "Why be Consistent? A Critical Analysis of Consistency Requirements in Choice Theory". *Economica* 52, pp. 167–183.

Sutherland, W. A. (1975). *Introduction to metric and topological spaces*. Oxford University Press.

Talagrand, O. (1997). "Assimilation of Observations, an Introduction". *Journal of the Meteorological Society of Japan* 75, pp. 191–209.

Taylor, B. N. and C. E. Kuyatt (1997). *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. Tech. rep. National Institute for Standards and Technology.

Taylor, T., Metroeconomica Ltd., and University of Bath (n.d.). *Heritage Building Case Study*. `http://www.ukcip.org.uk/wordpress/wp-content/PDFs/Costings/cost_cs_heritage.pdf`.

Tebaldi, C. and R. Knutti (2007). "The use of the multi-model ensemble in probabilistic climate projections". *Philosophical Transactions of the Royal Society* 365, pp. 2053–2075.

Topey, B. (2012). "Coin flips, credences and the Reflection Principle". *Analysis* 72, pp. 478–488.

Tversky, A. and D. Kahneman (1983). "Extension versus intuitive reasoning: The conjunction fallacy in probability judgement". *Psychological Review* 90, pp. 293–315.

——— (1986). "Rational Choice and the Framing of Decisions". *The Journal of Business* 59, S251–S278.

Uffink, J. (2007). "Compendium of the foundations of statistical mechanics". In: *Philosophy of Physics*. Ed. by J. Butterfield and J. Earman. North-Holland, pp. 923–1047.

van Fraassen, B. (1984). "Belief and the Will". *Journal of Philosophy* 81, pp. 235–256.

——— (1990). "Figures in a Probability Landscape". In: *Truth or Consequences*. Ed. by M. Dunn and K. Segerberg. Kluwer, pp. 345–356.

Van Horn, K. (2003). "Constructing a logic of plausible inference: a guide to Cox's theorem". *International Journal of Approximate Reasoning* 34, pp. 3–24.

Villegas, C. (1964). "On Qualitative Probability σ-algebras". *Annals of Mathematical Statistics* 35, pp. 1787–1796.

von Neumann, J. and O. Morgenstern ([1944] 2004). *Theory of Games and Economic Behaviour*. Sixtieth Anniversary Edition. Princeton University Press.

Walker, W., P. Harremoës, J. Rotmans, P. J. van der Sluijs, M. van Asselt, P. Janssen, and M. P. K. von Krauss (2003). "Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support". *Integrated Assessment* 4, pp. 5–17.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Vol. 42. Monographs on Statistics and Applied Probability. Chapman and Hall.

——— (2000). "Towards a unified theory of imprecise probabilities". *International Journal of Approximate Reasoning* 24, pp. 125–148.

Walley, P. and T. Fine (1982). "Towards a frequentist theory of upper and lower probability". *The Annals of Statistics* 10, pp. 741–761.

Weart, S. (2010). "The development of general circulation models of the climate". *Studies in the History and Philosophy of Modern Physics* 41, pp. 208–217.

Weatherson, B. (ms.). "Decision Making with Imprecise Probabilities". Available `http://brian.weatherson.org/vdt.pdf`.

Weisberg, J. (2009). "Commutativity or holism: A dilemma for conditionalizers". *British Journal for the Philosophy of Science* 60, pp. 793–812.

——— (2011). "Varieties of Bayesianism". In: *Handbook of the History of Logic*. Ed. by D. Gabbay, J. Woods, and S. Hartmann. Vol. 10. Elsevier.

Weisberg, M. (2006). "Robustness Analysis". *Philosophy of Science* 73, pp. 730–742.

Weitzman, M. L. (2009). *Some Basic Economics of Extreme Climate Change*.

Werndl, C. (2009). "What are the new implications of chaos for unpredictability?" *British Journal for the Philosophy of Science* 60, pp. 195–220.

Wheeler, G. (ms.). *Demystifying Dilation*.

Wheeler, G. and J. Williamson (2011). "Evidential Probability and Objective Bayesian Epistemology". In: *Philosophy of Statistics*. Ed. by P. S. Bandyopadhyay and M. Forster. Handbook of the Philosophy of Science. North-Holland, pp. 307–332.

White, R. (2010). "Evidential Symmetry and Mushy Credence". *Oxford Studies in Epistemology*.

Williams, J. R. G. (2012a). "Generalised Probabilism: Dutch Books and Accuracy Domination". *Journal of Philosophical Logic*.

——— (2012b). "Gradational accuracy and non-classical semantics". *Review of Symbolic Logic*.

Williamson, J. (2001). "Probability logic". In: *Handbook of the Logic of Inference and Argument: The Turn Toward the Practical*. Ed. by D. Gabbay, R. Johnson, H. Olbach, and J. Woods. North-Holland.

——— (2010). *In Defense of Objective Bayesianism*. Oxford University Press.

Willows, R. and R. Connell (2003). *Climate Adaptation: Risk, Uncertainty and Decision Making*. Tech. rep. http://www.ukcip.org.uk/wordpress/wp-content/PDFs/Risk.pdf. UKCIP.

Wilson, N. (2001). "Modified upper and lower probabilities based on imprecise likelihoods". In: *Proceedings of the 2nd International Symposium on Imprecise Probabilities and their Applications*.

Zynda, L. (2000). "Representation Theorems and Realism about Degrees of Belief". *Philosophy of Science* 67, pp. 45–69.