THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Bayesian Variations

Essays on the Structure, Object, and Dynamics of Credence

ARON VALLINDER

A thesis submitted to the Department of Philosophy, Logic and Scientific Method of the London School of Economics and Political Science for the degree of Doctor of Philosophy, London, 1 October 2018

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).

I confirm that Chapter 2 is based on a publication in *The British Journal for The Philosophy of Science* (Vallinder, 2018).

I declare that my thesis consists of 60,043 words.

Aron Vallinder

ABSTRACT

According to the traditional Bayesian view of credence, its structure is that of precise probability, its objects are descriptive propositions about the empirical world, and its dynamics are given by (Jeffrey) conditionalization. Each of the three essays that make up this thesis deals with a different variation on this traditional picture.

The first variation replaces precise probability with sets of probabilities. The resulting imprecise Bayesianism is sometimes motivated on the grounds that our beliefs should not be more precise than the evidence calls for. One known problem for this evidentially motivated imprecise view is that in certain cases, our imprecise credence in a particular proposition will remain the same no matter how much evidence we receive. In the first essay I argue that the problem is much more general than has been appreciated so far, and that it's difficult to avoid without compromising the initial evidentialist motivation.

The second variation replaces descriptive claims with moral claims as the objects of credence. I consider three standard arguments for probabilism with respect to descriptive uncertainty—representation theorem arguments, Dutch book arguments, and accuracy arguments—in order to examine whether such arguments can also be used to establish probabilism with respect to moral uncertainty. In the second essay, I argue that by and large they can, with some caveats. First, I don't examine whether these arguments can be given sound non-cognitivist readings, and any conclusions therefore only hold conditional on cognitivism. Second, decision-theoretic representation theorems are found to be less convincing in the moral case, because there they implausibly commit us to thinking that intertheoretic comparisons of value are always possible. Third and finally, certain considerations may lead one to think that imprecise probabilism provides a more plausible model of moral epistemology.

The third variation considers whether, in addition to (Jeffrey) conditionalization, agents may also change their minds by becoming aware of propositions they had not previously entertained, and therefore not previously assigned any probability. More specifically, I argue that if we wish to make room for reflective equilibrium in a probabilistic moral epistemology, we must allow for awareness growth. In the third essay, I sketch the outline of such a Bayesian account of reflective equilibrium. Given that (i) this account gives a central place to awareness growth, and that (ii) the rationality constraints on belief change by awareness growth are much weaker than those on belief change by (Jeffrey) conditionalization, it follows that the rationality constraints on the credences of agents who are seeking reflective equilibrium are correspondingly weaker.

ACKNOWLEDGMENTS

First of all, I owe tremendous thanks to my supervisors Richard Bradley and Anna Mahtani for their unfailing guidance, encouragement, and patience with my often half-baked ideas.

The LSE Philosophy Department has been a wonderful place in which to pursue this research. I'm grateful to my fellow PhD students for making these four years so much fun.

This thesis was written across four continents. I'm grateful to Haim Gaifman for hosting me at Columbia University, to Katie Steele for inviting me to the Australian National University, and to Ittay Nissan-Rozen and Ryan Doody for organising the 2018 Summer Workshop on Ethics and Uncertainty at the Hebrew University of Jerusalem.

I have received helpful comments from audiences at the 2016 Pitt-CMU Graduate Conference, the LSE Choice Group, the Higher Seminar in Theoretical Philosophy at Lund University, the 2016 London Intercollegiate Graduate Conference, and the 2018 Summer Workshop on Ethics and Uncertainty at the Hebrew University of Jerusalem. Chapter 2 was much improved by the suggestions of two anonymous referees for *The British Journal for the Philosophy of Science*.

For incisive comments and stimulating conversations, I'm grateful to Goreti Faria, Jim Joyce, Jurgis Karpus, David Kinney, Todd Karhu, Silvia Milano, Michael Nielsen, James Nguyen, Katie Steele, Bastian Stern, Reuben Stern, Rush Stewart, and Pablo Zendejas Medina.

My greatest thanks are to my parents Eva and Michael and my brother Jack for their love and support. Most importantly, I thank Elin, certainty among uncertainties.

CONTENTS

LIST OF TABLES

For this is action, this not being sure

— John Ashbery

Part I

THEME

1

BAYESIANISM

1.1 INTRODUCTION

Will it rain tomorrow? Is humanity still going to be around in a thousand years? Will global temperatures rise by more than 0.5°C over the next decade? Is there intelligent life elsewhere in the observable universe? Am I going to get to the station in time to catch the last tube home?

For these questions and many others, uncertainty seems like a sensible response. The world is complicated and our evidence limited. It would therefore seem foolish to invest all of one's confidence in a particular answer to one of these questions.

How should we reason when we are uncertain? We can distinguish between normative theories of reasoning under uncertainty by their different answers to the following questions:

(1) STRUCTURE How does it represent the agent's doxastic attitude?

(2) OBJECT What are the objects of uncertainty? What are we uncertain *about*?

(3) DYNAMICS How should our uncertainty change as we receive new evidence?

The most influential normative theory of reasoning under uncertainty has been the Bayesian theory. Traditional Bayesianism, as I will understand it, provides the following answers to our three questions: (1) with a single probability function; (2) empirical states of the world; and (3) in accordance with (Jeffrey) conditionalization. Traditional Bayesianism provides the theme for this thesis, the starting point for our investigations. Each essay explores the consequences of giving a different answer to one or more of these questions. As such, they constitute Bayesian variations, in the musical sense of the term: they are "version[s] of a theme, modified in melody, rhythm, harmony, or ornamentation, so as to present it in a new but still recognizable form" (*Oxford Dictionaries*).

The first essay, "Imprecise Bayesianism and Global Belief Inertia," considers a variation in the structure of credence: the so-called imprecise Bayesianism which models a rational agent's doxastic attitude with a set of probability functions instead of a single one. More specifically, it appraises a difficulty that arises for a particular evidentialist way of motivating imprecise credences. The second essay, "Moral Uncertainty and Arguments for Probabilism," examines a variation in the object of credence, and asks whether probabilism can

also be expected to hold when the domain of uncertainty is moral rather than descriptive. Finally, the third essay, "Bayesian Moral Epistemology and Reflective Equilibrium," considers a variation in the dynamics of credence, asking whether a Bayesian moral epistemology can accommodate reflective equilibrium. I argue that it can, provided that we depart from traditional Bayesianism by allowing for awareness growth, i.e. allowing for agents to become aware of propositions they had never considered or entertain before. Given that the ordinary Bayesian update rules are not applicable here, this raises the question of what, if any, constraints there are on the credences of agents who undergo awareness growth.

Although the three essays each deal with somewhat independent topics, some recurring themes and general lessons nevertheless emerge. These are summarised in the concluding chapter. The remainder of this chapter lays out the traditional Bayesian approach to epistemology in a bit more detail. In the next section, I first situate Bayesianism within a broader class of models of belief. In 1.3 I present the two core norms of traditional Bayesianism. In 1.4, I discuss the charge that the traditional Bayesian theory is excessively subjective, and present some further proposed constraints on rational credence. Finally, in 1.5, I discuss practical applications of Bayesian epistemology in the form of Bayesian decision theory.

1.2 THE STRUCTURE OF BELIEF

When you are uncertain, there is something you are uncertain *about*. You may be uncertain about what the weather will be like tomorrow, about what will happen a thousand years hence, about what distant regions of the universe look like, and so forth. These objects of uncertainty are required to have a particular structure. At the most general level, we will model them as elements of a σ-algebra.

σ-ALGEBRA. A σ-algebra $\mathcal{F}$ on a set Ω is a collection of subsets of Ω which contains Ω and is closed under complements and countable unions:

1. $\Omega \in \mathcal{F}$.
2. If $A \in \mathcal{F}$, then $\Omega \setminus A \in \mathcal{F}$.
3. If $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The underlying set Ω is called the *sample space* and its elements are called *outcomes*. Elements of $\mathcal{F}$ are called *events*. For example, suppose we are modelling an agent's beliefs about the possible outcomes of an experiment, such as the throw of a die. We can then interpret the conditions on σ-algebras as follows. The first condition says she must be able to entertain the proposition that some outcome or other occurs. The second condition says that if she is able to entertain some proposition, she must also be able to entertain its negation. Finally, the third condition says that if an agent is able to entertain some number of propositions individually, she must also be able to entertain their disjunction. Closure under countable unions is only necessary if the algebra has a countably infinite number of elements. This can only happen if the sample space Ω

is itself infinite. But sometimes we will be working with a finite Ω. In that case, we only need to assume closure under finite union, which gives us an *algebra of sets*.

In explaining the conditions on a σ-algebra, I appealed to logical notions like negation and disjunction. But of course strictly speaking such notions are not well-defined, because the objects we're concerned with are sets and not propositions. But there is another way to model the objects of uncertainty on which such notions are well-defined.

BOOLEAN ALGEBRA. A Boolean algebra is a six-tuple $\langle \mathscr{L}, \wedge, \vee, \neg, \bot, \top \rangle$ where $\mathscr{L}$ is a set such that for any $A, B, C \in \mathscr{L}$:

1. IDENTITY $A \vee \bot = A$ and $A \wedge \top = A$.
2. COMMUTATIVITY $A \vee B = B \vee A$ and $A \wedge B = B \wedge A$.
3. DISTRIBUTIVITY $A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$ and $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$.
4. COMPLEMENTS $A \vee \neg A = \top$ and $A \wedge \neg A = \bot$.

We can now define an implication relation $\vDash$ on $\mathscr{L}$ as follows:

$$A \vDash B \Leftrightarrow A \vee B = B \Leftrightarrow A \wedge B = A.$$

Using this implication relation we can instead define a Boolean algebra simply as the pair $\langle \mathscr{L}, \vDash \rangle$. In what follows, when I give formal definitions I will do so in terms of σ-algebras. However, informally I will often refer to the elements of a σ-algebra as propositions or simply *claims*. Sometimes philosophers favour Boolean algebras over σ-algebras because the elements of $\mathscr{L}$ can naturally be interpreted as *sentences* rather than propositions, thereby allowing us to potentially get around Frege's puzzles and related issues. For example, given that "Hesperus" and "Phosphorus" refer to the same celestial body, standard possible worlds semantics will identify the proposition "Hesperus is not Phosphorus" with the empty set, i.e. the contradictory proposition. But someone who believes that Hesperus is not Phosphorus seems to believe something different from someone who believes that $2 + 2 = 5$. In a sentence-based approach, we can represent these possibilities with distinct sentences, thereby making it possible for agents to rationally take different attitudes towards them. However, such intensional aspects of belief will not play a central role in our investigations, and I will therefore mainly stick to the set-based approach.

For philosophers, it is perhaps natural to think of Ω as a set of possible worlds. For that makes the algebra $\mathcal{F}$ a collection of sets of possible worlds, and we can then understand those sets of possible worlds as propositions in accordance with the familiar possible worlds semantics. We can think of $\mathcal{F}$ as the set of propositions of which the agent is aware, and I will therefore sometimes refer to $\mathcal{F}$ as the agent's *awareness set*. To say that an agent is aware of a proposition is to say that she is able to entertain it, i.e. that she is in the position to have propositional attitudes towards it. Typically, not much is said about which possible worlds go into Ω. Are they meant to be genuinely possible

worlds, i.e. maximally specific ways the world might be? And if so, does Ω have to contain *all* such maximally specific ways the world might be? Furthermore, is it possible for an agent's state of awareness to change over time? We will return to questions about awareness in chapter 4.

We now turn to the three main types of belief models. These models differ in how much structure they ascribe to our doxastic judgments, i.e. in how fine-grained they make those judgments out to be.

1.2.1 *Categorical Belief*

First, we have the traditional notion of full, or categorical, belief. An agent's categorical belief state is modelled as a (*categorical*) *belief set* $\mathcal{B} \subseteq \mathcal{F}$. For any $A \in \mathcal{F}$, we say that she *believes* A iff $A \in \mathcal{B}$, *disbelieves* A iff $\neg A \in \mathcal{B}$, and *suspends judgment* about A otherwise. Categorical belief is therefore a very coarse-grained notion of belief, allowing only for three types of doxastic attitudes.

We can ask various questions about the rationality of belief sets. Plausibly, any rational belief set should be *consistent*. It will be helpful to spell out consistency and some other related notions in a bit more detail here.[1] Let $Cn(\mathcal{B})$ be the closure of her belief set under logical consequence. The belief set $\mathcal{B}$ is

- *strictly inconsistent* iff there is some $A \in \mathcal{F}$ such that $A, \neg A \in \mathcal{B}$.
- *logically non-omniscient* iff there is some $A \in Cn(\mathcal{B})$ such that $A \notin \mathcal{B}$.
- *implicitly inconsistent* iff there is some A such that $A, \neg A \in Cn(\mathcal{B})$.

For example, I am strictly inconsistent if I believe both that it's going to rain tomorrow and that it's not going to rain tomorrow. I am logically non-omniscient if I believe both that it's going to rain tomorrow and that if it rains then I won't go for a run tomorrow, but I don't believe that I won't go for a run tomorrow. And I am implicitly inconsistent if I again believe both that it's going to rain tomorrow and that if it rains then I won't go for a run tomorrow, but I also believe that I will in fact go for a run tomorrow.

Let us introduce a bit more terminology. Say that a subset $\mathcal{S}$ of $\mathcal{F}$ is *maximal* just in case for every $A \in \mathcal{F}$, either $A \in \mathcal{S}$ or $\neg A \in \mathcal{S}$ (or both). And let a *completion* $\mathcal{C}$ of $\mathcal{B}$ be a maximal subset of $\mathcal{F}$ such that if $A \in \mathcal{B}$ then $A \in \mathcal{C}$. We can now say that an agent's beliefs are *coherently extendable* just in case there is some strictly consistent completion of her belief set. When an agent's beliefs are coherently extendable, this means that she could become fully opinionated (in the sense of coming to believe either A or $\neg A$ for every $A \in \mathcal{F}$) while holding on to all of her current beliefs, and remain consistent.

1 These definitions are taken from Bradley (2017:219–223).

1.2.2 *Relational Belief*

Some of our beliefs are more fine-grained than the categorical model allows. For example, I believe that it's more likely to rain tomorrow than it is to snow, and I take either of these two scenarios to be more likely than a hurricane. These are examples of *relational* belief. The basic judgments of relational belief are *comparative confidence judgments*, such as the judgment that A is at least as likely as B. We will let $A \trianglerighteq B$ represent this judgment. We model a relational doxastic state as a *relational belief set* $\mathcal{B}_{\trianglerighteq} \subseteq \mathcal{F} \times \mathcal{F}$ containing all comparative confidence judgments she accepts.

This means that we can define all the same consistency concepts as before. However, in order to do so we must first specify what notion of consequence the closure operator $Cn_{\trianglerighteq}$ appeals to. In effect, the consequence operator will be provided by the axioms that we settle on for the comparative confidence relation. Hence for any relational belief set $\mathcal{B}_{\trianglerighteq}$, $Cn(\mathcal{B}_{\trianglerighteq})$ will be its closure under comparative confidence consequence. For example, you might think that comparative confidence relation should be transitive, i.e. that if I believe that rain is more likely than snow and that snow is more likely than a hurricane, then I should also believe that rain is more likely than a hurricane. We will return to the question of which axioms to adopt for the comparative confidence relation in section 3.4.2.

1.2.3 *Quantitative Belief*

However, it seems that some of our beliefs are still more fine-grained than is allowed for by the relational model. For example, I might believe three times more strongly that it will rain tomorrow than that it will snow. To capture beliefs of this sort, we need a notion of *quantitative* belief. The basic judgments of quantitative belief are judgments of the form: "I believe A to degree x." We model a quantitative doxastic state as a *quantitative belief set* $\mathcal{B}_Q \subseteq \mathcal{F} \times \mathbb{R}$.

Again, we can define all the same consistency concepts, only now with respect to quantitative belief consequence. In the approach to quantitative belief with which we shall chiefly be concerned—the Bayesian one—this amounts to probabilistic consequence, i.e. consequence with respect to the probability axioms. For example, if I have 0.3 credence that it will rain tomorrow and 0.1 credence that it will snow, then it follows by probabilistic consequence that I should also have 0.4 credence that it will either rain or snow (assuming that it cannot do both).

1.2.4 *How the Models Relate to One Another*

The relational model is the most general, as it encompasses both the categorical and the quantitative as special cases. The categorical model can be viewed as a relational model that only distinguishes between two levels (or perhaps three, depending on how we treat suspension of judgment). Similarly, the probabilistic model can be viewed as a relational model that recognizes as many levels

as there are real numbers. Although the categorical model is useful for many purposes, for many others it is too coarse-grained. On the other hand, it seems the quantitative model may sometimes be too fine-grained. Perhaps relational belief can provide a happy middle ground. As we shall see, coherent extendability will have an important role to play in showing how relational belief and quantitative belief should hang together. But we are getting ahead of ourselves. Let us now turn our attention to the particular model of quantitative belief that we shall be concerned with, namely the probabilistic model.

1.3 THE TWO CORE BAYESIAN NORMS

1.3.1 *Probabilism*

Anyone familiar with games of chance will have some fluency in probabilistic reasoning. Suppose that you are about to throw a six-sided die. We let each side be an outcome, so that the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. When the sample space is so small, it is easy to let the algebra be the whole power set, i.e. $\mathcal{F} = 2^{\Omega}$. This means that every subset of Ω is a proposition that you can entertain and assign probability to. For example, $\{1\}$ will be the proposition that the die comes up 1, $\{1, 2\}$ will be the proposition that it comes up either 1 or 2, $\{2, 4, 6\}$ will be the proposition that it comes up even, etc.

Suppose we adopt the convention that the minimum degree of confidence is 0, and the maximum degree of confidence is 1. How should you then distribute your confidence among the possibilities? First of all, given that 0 is assumed to be the minimum degree of confidence, clearly no claim should receive confidence smaller than this. Second, we know that some side or other must come up, and hence the sample space Ω itself should receive maximum confidence. Finally, if two events are disjoint—that is, if they both cannot occur at once, such as the events $\{1\}$ and $\{2\}$—then one's confidence that one or the other will occur should be the sum of the degrees of confidence one assigns to them individually. In reasoning about this case, we have more or less stated the axioms on a probability function.

PROBABILITY FUNCTION A probability function on $\langle \Omega, \mathcal{F} \rangle$ is a function $P : \mathcal{F} \mapsto \mathbb{R}$ such that:

- P1. NON-NEGATIVITY $P(A) \geq 0$ for each $A \in \mathcal{F}$.
- P2. NORMALIZATION $P(\Omega) = 1$.
- P3. FINITE ADDITIVITY If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$.

The ordered triple $\langle \Omega, \mathcal{F}, P \rangle$ is called a *probability space*. The first axiom says that every outcome must receive probability greater than or equal to zero. The second axiom says that the probability that some number or other comes up is 1. The third axiom says that if two outcomes are disjoint—that is, if they cannot both occur at once—then the probability that one or the other occurs is

just the sum of their individual probabilities.[2] Let us briefly review some of the main consequences of the probability axioms:

1. $P(\emptyset) = 0$.
2. $P(A) \leq 1$ for each $A \in \mathcal{F}$.
3. If $\mathcal{A} = \{A_1, \ldots, A_n\}$ is a subset of $\mathcal{F}$ such that $A_i \cap A_j = \emptyset$ for each $A_i, A_j \in \mathcal{A}$, then $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.

These should all be fairly intuitive. The first says that the empty set—which we can think of as the contradictory proposition—should receive probability zero. This follows from P2 and P3. The second says that no proposition receives probability greater than 1. The third generalises P3. It says that for any number of mutually disjoint propositions, the probability of their union should be equal to the sum of the individual probabilities. For example, in the case of the six-sided die this means that $P(\{1,2,3\}) = P(\{1\}) + P(\{2\}) + P(\{3\})$.

With the notion of a probability function in place, we can now state the first normative claim of traditional Bayesianism:

PROBABILISM A rational agent's quantitative belief state can be represented as a probability space $\langle \Omega, \mathcal{F}, P \rangle$.

Probabilism thus makes demands both on how the agent represents the possibilities (namely that these form a σ-algebra) and on her doxastic attitudes towards these possibilities (namely that they satisfy the probability axioms). In many philosophical treatments of probabilism, the second demand is usually given more space than the first, but I will argue that scrupulous probabilists should also pay attention to the demand on representational capacities. This will be most evident in chapter 4, where I discuss the phenomenon of awareness growth.

1.3.2 *Conditionalization*

Often, we are interested in the probability of one event given that some other event has already occurred. For example, what is the probability that the die came up 1 given that it came up odd? To answer this, we must introduce the notion of conditional probability.

CONDITIONAL PROBABILITY If $P(B) > 0$, the probability of A *conditional* on B is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

The second core normative claim of traditional Bayesianism specifies how agents should revise their probabilistic beliefs as they receive more information:

2 Sometimes P3 is replaced by *countable* additivity: If $A_i \cap A_j = \emptyset$, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. See e.g. Easwaran (2013) for further discussion.

CONDITIONALIZATION If a rational agent with probability function $P(\cdot)$ learns proposition E with certainty and nothing else, her new probability function is given as $P_E(\cdot) = P(\cdot \mid E)$

Probabilism is a synchronic norm: it concerns the agent's degrees of belief at a given point in time. Conditionalization, on the other hand, is a diachronic norm: it concerns the relation between her degrees of belief at one point in time and her degrees of belief at another point in time. As the formulation above makes clear, conditionalization is only applicable when the learning experience takes the form of some definite proposition learned with certainty.

Some have claimed that all learning experiences are of this form, but this much more contentious claim is by no means a necessary component of the Bayesian picture. On the contrary, I believe Bayesians should recognise that there are many kinds of learning experience and that some of those kinds may require responses different from Conditionalization. Of particular interest is *uncertain learning*, i.e. learning without there being some proposition of which the agent becomes certain. When the agent learns some proposition E with certainty, what happens is that all probability is shifted to the first element of the partition $\{E, \neg E\}$. One way to generalise this is as follows. Let $\{B_i\}_{i=1}^n \subset \mathcal{F}$ be a partition of Ω.[3] Now let the learning experience have the effect of shifting, for each B_i her old probability assignment $P_1(B_i)$ to a new one, $P_2(B_i)$, which need not be extremal.

How should the agent revise her other degrees of belief, i.e. those not included in the partition, in light of this shift? Jeffrey conditionalization provides an answer:

JEFFREY CONDITIONALIZATION If a rational agent with probability function $P_1(\cdot)$ undergoes an uncertain learning experience which has the effect of shifting, for each element of some partition $\{B_i\}_{i=1}^n \subset \mathcal{F}$ of Ω, her probability assignment from $P_1(B_i)$ to $P_2(B_i)$, then for any proposition $A \in \mathcal{F}$ her new credence is given as:

$$P_2(A) = \sum_{i=1}^{n} P_1(A \mid B_i) \cdot P_2(B_i)$$

.

It is easy to see that ordinary conditionalization is a special case of Jeffrey conditionalization. Let our partition be $\{B, \neg B\}$, and let $P_2(B) = 1$. The formula for Jeffrey conditionalization then gives us

$$\begin{aligned} P_2(A) &= P_1(A \mid B)P_2(A) + P_1(A \mid \neg B)P_2(\neg B) \\ &= P_1(A \mid B) \cdot 1 + P_1(A \mid \neg B) \cdot 0 \\ &= P_1(A \mid B). \end{aligned}$$

Jeffrey conditionalization is appropriate when we have learned something without being able to specify exactly what it is we have learned. Consider Jeffrey's (1983:165) own original example:

3 That is, the elements are mutually exclusive (for any $i \neq j, B_i \cap B_j = \emptyset$) and jointly exhaustive ($\cup_{i=1}^n B_i = \Omega$).

> The agent inspects a piece of cloth by candlelight, and gets the impression that it is green, although he concedes that it might be blue or even (but very improbably) violet.

In better light conditions, the agent would perhaps be able to conditionalize on some proposition about the colour of the piece of cloth. But now that she is only able to get a faint impression, there does not seem to be any proposition available for her to conditionalize on. Hence Jeffrey conditionalization seems more appropriate: perhaps she assigns 0.8 credence to it being green, 0.19 to blue, and 0.01 credence to it being violet.

To better understand the relationship between conditionalization and Jeffrey conditionalization, we can break down conditionalization into two components. For any $E, H \in \mathcal{F}$:

RIGIDITY $P_E(H \mid E) = P(H \mid E)$

CERTAINTY $P_E(E) = 1$

Intuitively, Rigidity says that learning that E should not change how strongly you take E to support H. Certainty says that you must assign credence 1 to any proposition you have learned. Jeffrey conditionalization preserves Rigidity, but of course it violates Certainty.[4]

In summary, traditional Bayesianism is characterized by probabilism and (Jeffrey) conditionalization. However, that still leaves a lot to settle.[5] It will therefore be useful to survey some disputes within traditional Bayesianism, particularly as they will become relevant later on. First, we shall investigate whether there are any further norms on rational credences, beyond probabilism and conditionalization.

1.4 SUBJECTIVE VERSUS OBJECTIVE

In order for the Bayesian machinery to get going, we must have some probability function in place to begin with. If we don't have any values for $P(A)$ and $P(B)$, the conditional probability $P(A \mid B)$ will also be undefined and we will be unable to update on the evidence we receive. Of course, the values we assign to $P(A)$ and $P(B)$ will usually be based in part on evidence we have received in the past. But in order to have been able to update on that past evidence, we must already have had some probability function in place. The general point is this: no body of evidence (in the sense of a set of propositions that are assigned probability 1) is sufficient to determine a unique probability function.[6]

4 For more details on Jeffrey conditionalization, see Diaconis and Zabell (1982).

5 Good (1971) famously quipped that there are 46,656 varieties of Bayesians.

6 Except in the trivial case when the body of evidence includes, for every proposition in the algebra, either that proposition or its negation.

To put the point vividly, we can imagine a "superbaby" who is logically omniscient but lacks all empirical evidence.[7] Before she begins receiving any empirical evidence, the superbaby must somehow assign probability to all propositions of interest. These are her *prior probabilities*. Of course, you and I did not start our epistemic lives as such superbabies. However, consider an agent who satisfies probabilism and updates by standard conditionalization. For such an agent, their body of evidence $\mathcal{E} = \{E \in \mathcal{F} : P(E) = 1\}$ is the set of all propositions to which they assign credence one. As I will discuss in section 2.5, for any probability function P and body of evidence $\mathcal{E}$, there is a prior probability function P_0 such that $P(\cdot) = P_0(\cdot \mid \cap\mathcal{E})$. This means that any Bayesian agent can be represented as having some prior probability function, and we can ask what if any rationality constraints there are on priors.

This question of priors is an instance of the general debate over *uniqueness*, i.e. over whether for any given body of evidence there is a unique rational doxastic attitude to adopt in light of that evidence (White 2005). When the doxastic attitude in question is probabilistic, the question becomes a question of priors: in order for there to be a unique rational probability function to adopt in response to any body of evidence there would have to be a unique rational prior probability function.

Probabilism and conditionalization still leave a significant amount of freedom in the choice of credence function. The credence function has to satisfy the probability axioms and be updated in accordance with conditionalization. But how are we to choose among the vast number of probabilistic credence functions? According to *strict subjectivism*, this is the end of the matter as far as rationality is concerned: every probabilistic credence function is as good as any other. Some worry that without any further rationality constraints on prior probabilities, Bayesianism becomes a hopelessly subjective theory. Two agents may both be perfectly rational and receive the same evidence yet draw vastly different conclusions. Insofar as we find this worrying, there are two main kinds of responses. First, we can impose further rationality constraints on prior probabilities, so that the frequency of rational disagreement is reduced to an acceptable level. Second, we can argue that even if different agents start out with vastly different priors, they will eventually—assuming they both obtain the same veridical evidence—converge on the same opinion. Let us begin with the first strategy.

At the other end of the spectrum from strict subjectivism, we find *fully objective* Bayesians who believe that there is a unique rational prior probability function. As a result, they believe that for any body of evidence, there is a unique rational posterior probability function to adopt: namely the one that would result from conditionalising the unique rational prior on that particular body of evidence. However, I take it that most Bayesians fall somewhere in between these two extremes: they accept some further constraints on rational belief, without going so far as to say that there is always a uniquely most rational probability

7 Hájek (2012:418) attributes the term to David Lewis.

function to adopt. We shall now briefly survey proposed constraints that will be relevant for us later on.

1.4.1 *Regularity*

The non-negativity axiom requires that $P(A) \geq 0$ for every $A \in \mathcal{F}$. But consider now the following strengthening:

REGULARITY PRINCIPLE. For all $A \in \mathcal{F}$, if A is possible, then any rational prior probability function P_0 on $\mathcal{F}$ is such that $P_0(A) > 0$.

This principle entails that if $P_0(A) = 1$, then A is necessary. So a regular probability function will assign zero to all impossible propositions, one to all necessary propositions, and some number strictly in between to any contingent proposition. That sounds like a plausible constraint on prior probability functions: in particular, it can be seen as a way of being open-minded. Suppose that I initially assign credence zero to A. This means that no matter what evidence I then go on to receive, if I update by standard conditionalization, I will never be able to revise my credence in A one bit: if $P_0(A) = 0$ then $P_0(A \mid B) = 0$ whenever it is defined. Nor will I ever be able to conditionalize on A directly, because $P_0(B \mid A)$ will be undefined for every $B \in \mathcal{F}$.

So an irregular prior probability function forces the agent to forever remain probabilistically certain that some propositions are false come what may. Moreover, it does so on the basis of no evidence whatsoever. Therefore, we should ensure that agents begin their inquiries with an open mind by requiring their prior probability functions to be regular. 0 and 1 are rather different from all other probability assignments, because once a classical Bayesian agent has assigned probability 0 or 1 to some proposition, it is impossible for her to ever change her mind.

Of course, in order to evaluate this principle we must first specify which type of modality we are concerned with. There have been formulations in terms of logical possibility (Shimony 1955, Skyrms 1995), and metaphysical possibility (Lewis 1980). But perhaps the most plausible choice is *doxastic possibility* (Hájek, manuscript). A proposition is doxastically possible for an agent if it is logically consistent with what she believes. We can make this notion more precise as follows: a proposition A is doxastically possible for an agent just in case it is a non-empty element of her algebra $\mathcal{F}$. This gives us:

DOXASTIC REGULARITY PRINCIPLE. For any rational prior probability function P_0 on $\mathcal{F}$ and any non-empty $A \in \mathcal{F}$, $P_0(A) > 0$.

Suppose we are throwing darts at the unit interval, so that $\Omega = [0, 1]$. And suppose we let the σ-algebra simply be the power set of the sample space: $\mathcal{F} = 2^{\Omega}$. There is no countably additive P such that $P(A) > 0$ for all $A \in \mathcal{F}$. Any probability function defined on an uncountable algebra will be irregular. Hájek suggests that the the failure of regularity means that the ratio analysis will not do as a definition of conditional probability. According to that analysis, the

conditional probability $P(A \mid B)$ is only defined when $P(B) > 0$. As we have seen, it is impossible to assign non-zero credence to each of an uncountable number of propositions. Nevertheless, it still seems intelligible to speak of the probability conditional on such a proposition.

For example, suppose we are now throwing darts on a two-dimensional board, specified by Cartesian coordinates (x, y). It seems perfectly sensible to say that $P(y \geq 0 \mid x = 0) = 1/2$. That is, conditional on hitting the board at $x = 0$, the probability of hitting above the origin is one half. But of course $P(x = 0) = 0$, so the ratio analysis cannot make sense of this conditional probability. Nor can it make sense of the intuitive idea that the probability of any proposition *conditional on itself* should be 1: if $P(B) = 0$ then $P(B \mid B)$ will be undefined. Furthermore, if it is sometimes necessary to assign credence zero to doxastically possible propositions, then it may happen that an agent learns a proposition she previously assigned credence zero. What should she do then?

Conditionalization tells her to replace her old probabilities with conditional probabilities. But if we maintain the ratio analysis, these will be undefined when she learns a probability zero proposition. In response, some have suggested that we should take conditional probability to be the primitive notion and define unconditional probability in terms of conditional probability rather than the other way. This way we can ensure that the ratio formula still holds whenever the probability of the conditioning proposition is non-zero while also making room for conditional probability in cases where this condition is not met. We shall return to the regularity principle in section 2.5 when we discuss a problem for imprecise Bayesianism.

1.4.2 *Chance-Credence Principles*

Suppose we think there is such a thing as objective chance. Perhaps a coin being flipped has some genuine objective chance of coming up heads. Or perhaps there is objective chance at the quantum level.[8] If we do believe in objective chance, it's natural to think that knowledge of objective chance should influence our credences in some way. In particular, if you know what the objective chance of some event is, then it seems you should set your credence that the event will occur to equal the objective chance. Something like this idea seems to be at play in the way we usually think about coin flips: it is reasonable to assign probability one half to heads precisely because we take that to be the objective chance of heads. Various so-called *chance-credence principles* make this intuitive idea more precise. The most famous such proposal is the principal principle (Lewis 1980).

PRINCIPAL PRINCIPLE Let P_0 be any rational prior probability function. Let t_i be any time. Let $\mathrm{ch}_i(A) = x$ be the proposition that the chance at t_i

8 For an argument that objective chance at some level of description is compatible with determinism at a lower level of description, see List and Pivato (2015).

of A is x. Let E be any proposition consistent with $\mathrm{ch}_i(A) = x$ that is admissible at t_i. Then

$$P_0(A \mid \mathrm{ch}_i(A) = x \wedge E) = x.$$

To fully unpack this, we need to say what it is for a proposition to be admissible at a time. Among other things, Lewis intended for the criterion to rule out cases of foreknowledge: if you are somehow able to know, before the fact, that the coin will come up heads, then your credence in this proposition should clearly be 1, even if you know that its chance is 0.5. The details of this debate are not central to our discussion, although chance-credence principles will also become relevant in section 2.5 when we consider what constraints to impose on imprecise priors.

1.4.3 *Reflection Principle*

The principal principle is an example of a deference principle, or an expert principle: it tells us to defer to the expertise of objective chance. Another expert principle is van Fraassen's (1984) reflection principle:

REFLECTION PRINCIPLE If P_1 is the agent's credence function at some time t_1 and P_2 her credence function at some later time t_2, then for any $A \in \mathcal{F}$,

$$P_1(A \mid P_2(A) = x) = x.$$

Of course, we rarely get information about our future credences other than through the passage of time. And there are cases where the principle clearly shouldn't apply, at least as stated. For example, even if I know that when I get drunk tonight I will believe that I'm a fantastic driver, I should not now conclude that I will be a fantastic driver tonight. Even so, there may be some use for reflection. Here is how van Fraassen (1995:25–26) views the principle:

> Integrity requires me to express my commitment to proceed in what I now classify as a rational manner, to stand behind the ways in which I shall revise my values and opinions.

As we shall see in chapter 2, some object to imprecise Bayesianism on the grounds that it occasionally violates a generalised version of the reflection principle.

1.4.4 *Principle of Indifference*

Unlike the previous principles, the principle of indifference seeks to narrow down the range of rational prior probability functions to just a single one. It starts with the plausible-sounding idea that if one has no reason to think that A is more likely than B or vice versa, then one should treat them as equally likely; and it gives this idea a precise formulation: whenever there is some finite number of mutually exclusive and jointly exhaustive possibilities and one has no reason to think that any one of them is more likely than any other, then

one should assign all of them equal probability.[9] [10]

Various justifications of the principle of indifference have been proposed. White (2009) argues that the principle provides the correct way of responding to evidence. Jaynes (1957) argues that of all prior probability functions one might adopt, the indifference prior encodes the least information. Williamson (2017) argues that the indifference prior is in a certain sense the most cautious one. And Pettigrew (2016b) gives an accuracy-based argument.[11]

As is well-known, this principle yields different verdicts depending on how the possibilities are described. How likely is it to rain tomorrow? If the possibilities are rain and no rain, the probability is $1/2$. But if the possibilities are moderate rain, heavy rain, and no rain, the probability is $2/3$. So on the face of it, the principle of indifference appears to give inconsistent advice.

If we wish to defend the principle against the charge of inconsistency, we have two options: either claim that there is some privileged way of describing the possibilities and that the principle of indifference should only be applied to this description, or claim that the principle is a description-relative constraint which only comes into play once we've settled on a particular description. The first option avoids inconsistency because it takes the principle to only ever issue a unique recommendation, so there can never be multiple recommendations that may conflict with one another. The second option avoids inconsistency because it takes the principle to be applicable only after we've specified a description of the possibilities. Once we have such a description, the principle will issue a unique recommendation and the fact that it would issue different recommendations for other descriptions is neither here nor there.

Although there have been several ingenious attempts to show that many problem cases for the principle of indifference do in fact have a privileged description (e.g. Jaynes 1973), I take the current consensus to be that such a strategy is unlikely to succeed across the board. Indeed, among recent defenders of the principle, many have explicitly endorsed the second response to the charge of inconsistency.[12] If we go for the description-relative formulation, then the propositions that are to be assigned equal prior probability must be elements of the agent's algebra $\mathcal{F}$ that together form a partition of Ω. Of the various

9 Here I intentionally set aside issues that arise when the number of possibilities is either countably or uncountably infinite.

10 For example, the early Wittgenstein appears to have advocated a version of this view (cf. 5.15–5.154 of the *Tractatus*).

11 See Jaynes (2003) for an extended discussion and defense of the principle.

12 For example, Pettigrew (2016: 57), who offers an accuracy-based argument for indifference, writes that "the Principle of Indifference will make different demands on [an agent] depending on the set of propositions she entertains." Similarly, in his defence of the different but closely related Maximum Entropy Principle, Williamson (2010: 156) writes that "[t]here is no getting round it: probabilities generated by the Maximum Entropy Principle depend on language as well as evidence."

partitions available, I take it that the natural choice is the most fine-grained one.[13,14] This gives us the following:

PRINCIPLE OF INDIFFERENCE. A rational agent with algebra $\mathcal{F}$ over Ω should assign equal prior probability to each cell of the finest partition $\mathcal{A}$ of Ω such that $\mathcal{A} \subset \mathcal{F}$:

$$P_0(A) = \frac{1}{|\mathcal{A}|} \text{ for each } A \in \mathcal{A}.^{15}$$

The principle of indifference will be especially important in the next chapter. As we shall see, Joyce (2005, 2010) argues that even if we could formulate the principle in a way that avoided inconsistency it would still be objectionable, because it commits us to very definite beliefs when there is no evidence available to support such beliefs. This rejection of the principle of indifference forms part of the argument for his for the brand of evidentially-motivated imprecise Bayesianism we will examine in that chapter.

1.4.5 *Washing Out Theorems*

We have now surveyed some putative further constraints on rational probability functions. How worried should we be if it turns out that very few of these are acceptable? Would it make Bayesianism unacceptably subjective? Some think that so-called *washing out theorems* can still vindicate the objectivity of Bayesian epistemology. Roughly speaking, such theorems show that under certain conditions, agents who begin with different prior probability functions will, as they acquire more and more evidence, in the long run converge on the same posterior probability function. So although the choice of prior is subjective, as agents acquire more and more evidence their priors play a smaller and smaller part in determining their posteriors until they eventually converge on the same posterior. In the long run, therefore, the subjective aspect of Bayesianism is weeded out.

One influential convergence theorem is due to Blackwell and Dubins (1962). Consider two agents who assign probability to the set of all infinite binary sequence. At each step, they both observe and conditionalize on one new element of the sequence. Let their respective prior probability functions be P_1 and P_2. Say that P_1 is *absolutely continuous* with respect to P_2 just in case for any $A \in \mathcal{F}$, $P_1(A) > 0$ implies $P_2(A) > 0$. And say that two sequences of probability functions *merge* if the values they assign eventually stay within some ϵ of

13 Given that we are assuming that $\mathcal{F}$ is finite, there will always exist a unique most fine-grained partition $\mathcal{P}$ of Ω such that $\mathcal{P} \subset \mathcal{F}$, provided only that we exclude the trivial algebra $\mathcal{F} = \{\emptyset, \Omega\}$.

14 Most defenders of the description-relative principle of indifference also seem to endorse this way of applying it. For example, Pettigrew (2016: 57) writes that the principle of indifference requires of an agent that she "divide her credence equally over the possibilities *grained as finely as the propositions she entertains will allow*" (emphasis in original).

15 Given that $\mathcal{A}$ is the finest partition of Ω contained in $\mathcal{F}$, every element of $\mathcal{F}$ can be written as the union of some elements of $\mathcal{A}$ (all of which are disjoint), and hence PI determines a unique prior probability assignment for each element of $\mathcal{F}$.

one another. What Blackwell and Dubins found is that if P_1 is absolutely continuous with P_2, then they will eventually merge with P_1 probability 1. This means that if I assign probability zero to every proposition that you assign probability zero, then I will be certain (in the sense of assigning probability 1) that you and I will eventually merge. Correspondingly, if you also assign probability zero to every proposition that I assign probability zero, then you will be certain (in the sense of assigning probability 1) that you and I will eventually merge.

You may wonder, especially in light of the objections to the regularity principle that we discussed, whether it's plausible to require that the agents only assign probability zero to the same propositions. Or you may wonder about the fact that merging is only guaranteed in the sense that the agents themselves are certain that it will happen—couldn't they be mistaken? Finally, you may wonder what relevance results like these have for agents like ourselves, who will only ever receive finite amounts of evidence. Granted, it is better from the viewpoint of objectivity that this result is indeed a theorem than it would be were its negation a theorem. But just how much reassurance we derive from this result and others like it will depend on our assessment of their conditions.[16] This concludes our survey of subjectivism versus objectivism. Next, we turn to some practical uses for Bayesian epistemology.

1.5 DECISION THEORY

Although this thesis is primarily concerned with Bayesian epistemology and not Bayesian decision theory, the two often go hand in hand. And I take it that if you find the former compelling as an account of theoretical rationality, you will be predisposed to find the latter compelling as an account of practical rationality. Moreover, some very influential arguments for probabilism are in fact arguments for Bayesian decision theory—i.e., arguments for expected utility maximisation—of which probabilism is a consequence. Furthermore, the formal framework we shall use for representing moral claims draws heavily on some aspects of decision theory. Therefore a refresher may prove handy.[17] Decision theory starts with *decision problems*: situations in which an agent faces a choice between different options. For example, suppose you are considering whether or not to bring an umbrella as you leave the house. The relevant question, of course, is whether it's going to rain or not. We can represent this decision problem using a state-consequence matrix as follows:

So the picture is this: agent chooses between some number of options, and those options have certain consequences. Which particular consequence will follow from the exercise of a given option depends on the state of nature. Whether leaving the umbrella will have the consequence of getting soaked depends on whether it's raining or not. If you are certain that it will rain, the

16 Another important result is due to Gaifman and Snir (1982). See Earman (1992, chapter 6) and Hawthorne (2008) for further discussion.

17 See Jeffrey (1983), Joyce (1999) and Bradley (2017) for detailed accounts of Bayesian decision theory.

	Clear	*Rainy*
Umbrella	Dry, carrying umbrella	Dry, carrying umbrella
No umbrella	Dry, no umbrella	Wet, no umbrella

Table 1: *State-Consequence Matrix*

decision problem is easy: just bring the umbrella. Similarly, if you are certain that it won't rain, you'll know not to bother. But if you're uncertain, things aren't quite as straightforward. There's some chance that it will rain, in which case you would strongly prefer having the umbrella to not having it. But there's some chance that it won't, in which case it would be a minor nuisance to have to carry an umbrella around. How should you make up your mind?

Bayesian decision theory answers that you should maximise expected utility.[18] The core idea is that what you should do depends both on your level of uncertainty and how strongly you desire the different possible outcomes. First, we represent the agent's uncertainty as a probability function over states, in this case simply Rain and No Rain. Second, we use a utility function to represent how desirable the agent finds the various possible consequences. Given that having to carry an umbrella around on a cloudless day is mild inconvenience compared to the serious frustration of getting your new business suit soaked, this should be reflected in the numbers that the utility function assigns to these consequences.

	0.5	*0.5*
Umbrella	9	9
No umbrella	10	1

Table 2: *Probability-Utility Matrix*

We can now give expected utility maximization a rough first mathematical formulation. There are various ways of doing so, and I will be using the framework of Jeffrey (1965). The differences between this framework and others will not concern us here. The probability function P is defined as before. The utility function $U : \mathcal{F} \mapsto \mathbb{R}$ assigns a number to each possibility, indicating how desirable the agent finds that possibility. Given that we are uncertain about the consequences of our options, we can think of an option as a probability distribution over outcomes. So in deciding between her options, the agent is deciding between different probability distributions over consequences. Let $\mathcal{O} = \{O_1, \ldots, O_n\}$ be her options. Then for any proposition $A \in \mathcal{F}$, $P(A \mid O_i)$

18 The idea of expected utility maximisation is as old as probability theory itself. In the *Port-Royal Logic* we learn that "to judge what one ought to do to obtain a good or avoid an evil, one must not only consider the good and the evil in itself, but also the probability that it will or will not happen and view geometrically the proportion that all these things have together." (Arnauld and Nicole 1662/1996).

will be the probability of A conditional on having performed option O_i. The expected utility of a given option O can now be written as:

$$\mathbb{E}(O) = \sum_{A \in \mathcal{F}} U(A)P(A \mid O).$$

And the norm of expected utility maximisation now says that rationality requires agents to always choose an option whose expected utility is at least as high as that of any other available option.

1.6 OVERVIEW

Let us summarize. Recall the three questions with which we began:

(1) STRUCTURE How does the theory represent the agent's doxastic attitude?

(2) OBJECT What are the objects of uncertainty? What are we uncertain *about*?

(3) DYNAMICS How should our uncertainty change as we receive new evidence?

Traditional Bayesianism gives the following answers:

(1) The agent's doxastic attitude is represented as a probability space.

(2) The objects of credence are descriptive propositions.

(3) The agent should update her credences by (Jeffrey) conditionalization.

We have seen that this still leaves room for a lot of disagreement. Strict subjectivists believe that there are no other constraints on rational credence, whereas moderate subjectivists and full-blown objectivists accept further constraints, such as the regularity principle, chance-credence principles, the reflection principle, or the principle of indifference. And some think that convergence results can help us evade the charge of subjectivism even if there are only relatively few constraints on rational credence. We have also seen that Bayesian epistemology goes well together with Bayesian decision theory, according to which rational agents perform the action that maximises expected utility relative to their probability and utility functions.

In the next part of this thesis, we will consider three variations on these three answers. In chapter 2, we consider what happens when the structure of credence is given as a set of probability functions rather than a single one. In chapter 3 we examine whether probabilism can be justified when the objects of credence are moral claims rather than descriptive claims. And in chapter 4 we attempt to provide a Bayesian account of reflective equilibrium by allowing the agent's credences to change over time by awareness growth as well as by (Jeffrey) conditionalization. Although these three chapters concern somewhat different topics, some general lessons nevertheless emerge. These are summarised in chapter 5.

Part II

VARIATIONS

2

IMPRECISE BAYESIANISM AND GLOBAL BELIEF INERTIA

2.1 INTRODUCTION

Our first variation concerns the structure of credence. In the traditional Bayesian framework, agents must have precise degrees of belief, in the sense that these degrees of belief are represented by a real-valued credence function. This may seem implausible in several respects. In particular, one might think that our evidence is rarely rich enough to justify this kind of precision—choosing one number over another as our degree of belief will often be an arbitrary decision with no basis in the evidence. For this reason, Joyce (2010) suggests that we should represent degrees of belief by a set of credence functions instead.[1] This way, we can avoid arbitrariness by requiring that the set contains all credence functions that are, in some sense, compatible with the evidence.

However, this requirement creates a new difficulty. The more limited our evidence is, the greater the number of credence functions compatible with it will be. In certain cases, the number of compatible credence functions will be so vast that the range of our credence in some propositions will remain the same no matter how much evidence we subsequently go on to obtain. This is the problem of belief inertia. Joyce is willing to accept this implication, but I will argue that the phenomenon is much more widespread than he seems to realize, and that there is therefore decisive reason to abandon his view.

In the next section, I provide some reason for thinking that the precision of the traditional Bayesian framework may be problematic. In Section 3, I present Joyce's preferred alternative—imprecise Bayesianism—and attempt to spell out its underlying evidentialist motivation. In particular, I suggest an account of what it means for a credence function to be compatible with a body of evidence. After that, in Section 4, I introduce the problem of belief inertia via an example from Joyce. I also prove that one strategy for solving the problem (suggested but not endorsed by Joyce) is unsuccessful. Section 5 argues that the problem is far more general than one might think when considering Joyce's example in isolation. The argument turns on the question of what prior credal state an evidentially motivated imprecise Bayesian agent should have. I maintain that, in light of her motivation for rejecting precise Bayesianism, her prior credal state must include all credence functions that satisfy some very weak constraints. However, this means that the problem of belief inertia is with us

1 Although Joyce is my main target in this essay, the view is of course not original to him. For an influential early exponent, see Levi (1980).

from the very start, and that it affects almost all of our beliefs. Even those who are willing to concede certain instances of belief inertia should find this general version unacceptable. Finally, in section 6 I consider a few different ways for an imprecise Bayesian to respond. The upshot is that we must give up the very strong form of evidentialism and allow that the choice of prior credal state is to a large extent subjective.

2.2 THE PROBLEMS WITH PRECISION

As we saw in the previous chapter, traditional Bayesianism is committed to the following two normative claims:

PROBABILISM A rational agent's quantitative belief state can be represented as a probability space $\langle \Omega, \mathcal{F}, P \rangle$.

CONDITIONALIZATION If a rational agent with probability function $P_1(\cdot)$ learns proposition E with certainty and nothing else, her new probability function is given as $P_2(\cdot) = P_1(\cdot \mid E)$.

Some philosophers within the Bayesian tradition have taken issue with the precision required by probabilism. For one thing, it may appear descriptively inadequate. It seems implausible to think that flesh-and-blood human beings have such fine-grained degrees of belief.[2] However, even if this psychological obstacle could be overcome, Joyce (2010) argues that precise probabilism should be rejected on normative grounds, because our evidence is rarely rich enough to justify having precise credences. His point is perhaps best appreciated by way of example.

THREE URNS There are three urns in front of you, each of which contains a hundred marbles. You are told that the first urn contains fifty black and fifty white marbles, and that all marbles in the second urn are either black or white, but you don't know their ratio. You are given no further information about marble colours in the third urn. For each urn i, what credence should you have in the proposition B_i that a marble drawn at random from that urn will be black?

Here I will understand a random draw simply as one where each marble in the urn has an equal chance of being drawn. That makes the first case straightforward. We know that there are as many black marbles as there are white ones, and that each of them has an equal chance of being drawn. Hence we should apply some chance-credence principle and set $P(B_1) = 0.5$.[3] The second case is not so clear-cut. Some will say that any credence assignment is permissible, or at least that a wide range of them are. Others will again try to identify a unique credence assignment as rationally required, typically via an application of the principle of indifference. They will claim that we have no

2 Whether this is implausible will depend on what kind of descriptive claim one thinks is involved in ascribing a precise degree of belief to an agent. See for instance Meacham and Weisberg (2011).

3 Hardcore subjectivists may insist that, even in this case, any probabilistically coherent credence assignment is permissible.

reason to consider either black or white as more likely than the other, and that we should therefore give them equal consideration by setting $P(B_2) = 0.5$.

However, as is well-known, the principle of indifference gives inconsistent results depending on how we partition the space of possibilities.[4] This becomes even more evident when we consider the third urn. In the first two cases we knew that all marbles were either black or white, but now we don't even have that piece of information. So in order to apply the principle of indifference, we must first settle on a partition of the space of possible colours. If we settle on the partition $\{\text{black}, \text{not black}\}$, the principle of indifference gives us $P(B_3) = 0.5$. If we instead think that the partition is given by the eleven basic colour terms of the English language, the principle of indifference tells us to set $P(B_3) = 1/11$.

How can we determine which partition is appropriate? In some problem cases, the principle's adherents have come up with ingenious ways of identifying a privileged partition. However, Joyce (2005:170) argues that even if this could be done across the board (which seems doubtful), the real trouble runs deeper. The principle of indifference goes wrong by always assigning precise credences, and hence the real culprit is (precise) probabilism. In the first urn case, our evidence is rich enough to justify a precise credence of 0.5. But in the second and third cases, our evidence is so limited that any precise credence would constitute a leap far beyond the information available to us. Adopting a precise credence in these cases would amount to acting as if we have evidence we simply do not possess, regardless of whether that precise credence is based merely on personal opinion, or whether it has been derived from some supposedly objective principle.

The lesson Joyce draws from this example is therefore that we should only require agents to have imprecise credences. This way we can respect our evidence even when that evidence is ambiguous, partial, or otherwise limited. My target in this paper will be this sort of evidentially motivated imprecise Bayesianism. In the next section I present the view and clarify the evidentialist argument for adopting it.

2.3 IMPRECISE BAYESIANISM

Joyce's (2010:287) imprecise Bayesianism makes the following two normative claims:

IMPRECISE PROBABILISM A rational agent's quantitative beliefs can be represented as a triple $\langle \Omega, \mathcal{F}, \mathcal{P} \rangle$, where $\mathcal{P} = \{P_1, P_2, \dots\}$ is a set of probability functions on $\mathcal{F}$.

4 Widely discussed examples include Bertrand's (1889) paradox, and van Fraassen's (1989) cube factory.

IMPRECISE CONDITIONALIZATION. If a rational agent with credal state $\mathcal{P}$ learns proposition E with certainty and nothing else, her new credal state is given as $\mathcal{P}_E = \{P_i(\cdot \mid E) : P_i(\cdot) \in \mathcal{P}\}$.[5]

Each individual credence function thus behaves just like the credence functions of precise Bayesianism: they are probabilistic, and they are updated by conditionalization. The difference is only that the agent's degrees of belief are now represented by a set of credence functions, rather than a single one. As a useful terminological shorthand, I will write $\mathcal{P}(A)$ for the set of numbers assigned to the proposition A by the elements of $\mathcal{P}$, so that $\mathcal{P}(A) = \{x : \exists P \in \mathcal{P}$ s.t. $P(A) = x\}$. I will refer to $\mathcal{P}(A)$ simply as the agent's credence in A.

Agents with precise credences are more confident in a proposition A than in another proposition B if and only if their credence function assigns a greater value to A than to B. In order to be able to make similar comparisons for agents with imprecise credences, we will adopt what I take to be the standard, supervaluationist, view and say that an imprecise believer is determinately more confident in A than in B if and only if $P(A) > P(B)$ for each $P \in \mathcal{P}$. If there are $P_1, P_2 \in \mathcal{P}$ such that $P_1(A) > P_1(B)$ and $P_2(A) < P_2(B)$, it is indeterminate which of the two propositions she regards as more likely. In general, any claim about her overall doxastic state requires unanimity among all the credence functions in order to be determinately true or false.[6]

Imprecise Bayesianism has been objected to in several ways. In cases of so-called *dilation*, imprecise Bayesianism entails that a rational agent's credal state will foreseeably become less precise as she acquires new evidence (Seidenfeld and Wasserman 1993, Bradley and Steele 2014). Some find this objectionable, either because it seems to violate a reflection principle for imprecise credences, or because of a conviction that evidence should serve to make our credences more rather than less precise. Others have argued that any decision theory for imprecise probability is either implausible or collapses into a decision theory for precise probability, thereby calling into question the need for imprecision in the first place (see e.g. Elga 2010 and Mahtani 2018). However, the objection I will present is specific to an evidentially motivated imprecise Bayesianism. Let us therefore try to spell this out in a bit more detail.

Joyce defends imprecise Bayesianism on the grounds that many evidential situations do not warrant precise credences. With his framework in place, we can respect the datum that a precise credence of 0.5 is the correct response in the first urn case, without thereby being forced to assign precise credences in

5 As stated, the update rule doesn't tell us what to do if an element of the credal state assigns zero probability to a proposition that the agent later learns. This problem is of course familiar from the precise setting. Three options suggest themselves: (i) discard all such credence functions from the posterior credal state, (ii) require that each element of the credal state the regularity principle, so that they only assign zero to doxastically impossible propositions, thereby ensuring that the situation can never arise, or (iii) introduce a primitive notion of conditional probability. For my purposes, we don't need to settle on a solution. I'll just assume that the imprecise Bayesian has some satisfactory way of dealing with these cases.

6 This supervaluationist view of credal states is endorsed by Joyce (2010), van Fraassen (1990), and Hájek (2003), among others.

the second and third cases as well. In these last two cases, our evidence is ambiguous or partial, and assigning precise credences would require making a leap far beyond the information available to us.

This raises the question of how far in the direction of imprecision we should move in order to remain on the ground. How many credence functions must we include in our credal state before we can be said to be faithful to our evidence? Joyce answers that we should include just those credence functions that are compatible with our evidence.[7] We can state this as:

EVIDENCE GROUNDING THESIS At any point in time, a rational agent's credal state includes all and only those credence functions that are compatible with the total evidence she possesses at that time.

To unpack this principle, we need a substantive account of what it takes for a credence function to be compatible with a body of evidence. One such proposal is due to White (2010:174):

CHANCE GROUNDING THESIS Only on the basis of known chances can one legitimately have sharp credences. Otherwise one's spread of credence should cover the range of possible chance hypotheses left open by your evidence.

The chance grounding thesis posits a very tight connection between credence and chance. As Joyce (2010:289) points out, the connection is indeed too tight, in at least one respect. There are cases where all possible chance hypotheses are left open by our evidence, but where we should nevertheless have sharp (precise) credences. He provides the following example.

SYMMETRICAL BIASES Suppose that an urn contains coins of unknown bias, and that for each coin of bias α there is another coin of bias $(1-\alpha)$. One coin has been chosen from the urn at random. What credence should we have in the proposition H, that it will come up heads on the first flip?

Because the chance of heads corresponds to the bias of the chosen coin (whatever it is), and since (for all we know) the chosen coin could have any bias, every possible chance hypothesis is left open by the evidence. In this setup, for each $P \in \mathcal{P}$, the credence assignment $P(H)$ is given as the expected value of a corresponding probability density function (pdf), f_P, defined over the possible chance hypotheses: $P(H) = \int_0^1 x \cdot f_P(x)\, dx$. The information that, for any α, there are as many coins of bias α as there are coins of bias $(1-\alpha)$ translates into the requirement that for each $a, b \in [0,1]$ and for every f_P,

$$\int_a^b f_P(x)\, dx = \int_{1-b}^{1-a} f_P(x)\, dx. \tag{1}$$

Any f_P which satisfies this constraint will be symmetrical around the midpoint, and will therefore have an expected value of 0.5. This means that $P(H) = 0.5$

7 Joyce (2010:288) writes that each element of the credal state is a probability function that the agent takes to be compatible with her evidence. This formulation leaves it open whether compatibility is meant to be an objective or a subjective notion; we will return to this issue later.

for each $P \in \mathcal{P}$. Thus we have a case where all possible chance hypotheses are left open by the evidence, but where we should still have a precise credence.[8]

Nevertheless, something in the spirit of the chance grounding thesis looks like a natural way of unpacking the evidence grounding thesis. In Joyce's example, each possible chance hypothesis is indeed left open by the evidence, but we do know that every pdf f_P must satisfy constraint (1) for each $a, b \in [0, 1]$. So any f_P which doesn't satisfy this constraint will be incompatible with our evidence. And similarly for any other constraints our evidence might impose on f_P. In the case of a known chance hypothesis, the only pdf compatible with the evidence will be the one that assigns all weight to that known chance value. Similarly, if the chance value is known to lie within some particular range, then the only pdfs compatible with the evidence will be those that are equal to zero everywhere outside of that range.

However, as Joyce's example shows, these are not the only ways in which our evidence can rule out pdfs. More generally, evidence can constrain the shape of the compatible pdfs. In light of this, we can propose the following revision.

REVISED CHANCE GROUNDING THESIS A rational agent's credal state contains all and only those credence functions that are given as the expected value of some probability density function over chance hypotheses that satisfies the constraints imposed by her evidence.

Just like White's original chance grounding thesis, my revised formulation posits an extremely tight connection between credence and chance. For any given body of evidence, it leaves no freedom in the choice of which credence functions to include in one's credal state. Because of the way compatibility is understood, there will always be a fact of the matter about which credence functions are compatible with one's evidence, and hence about which credence functions ought to be included in one's credal state.

The question, then, is whether we should settle on this formulation, or whether we can change the requirements without thereby compromising the initial motivation for the imprecise model. In his discussion of the chance grounding thesis, Joyce (2010:288) claims that even when the error in White's formulation has been taken care of, as I proposed to do with my revision, the resulting principle is not essential to the imprecise proposal. Instead, he thinks it is merely the most extreme view an imprecise Bayesian might adopt. Now, this is certainly correct as a claim about imprecise Bayesianism in general. One can

8 It has been suggested to me that it might make a difference whether the coin that is to be flipped has been chosen yet or not. If it has not yet been chosen, a precise credence of 0.5 seems sensible in light of one's knowledge of the setup. If instead it has already been chosen, then it has a particular bias, and since the relevant symmetry considerations are no longer in play, one's credence should be maximally imprecise: $[0, 1]$. However, one might argue that rationally assigning a precise credence of 0.5 when the coin has not yet been chosen does not constitute a counterexample to the original chance grounding thesis, by arguing that the proposition 'The next coin to be flipped will come up heads' has an objective chance of 0.5. My argument won't turn on this, so I'm happy to go along with Joyce and accept that we have a counterexample to the chance grounding thesis.

accept both imprecise probabilism and imprecise conditionalization without accepting any claim about how knowledge of chance hypotheses, or any other kind of evidence, should constrain which credence functions are to be included in the credal state. However, on the evidentially motivated proposal that Joyce advocates himself, it's not clear whether any other way of specifying what it means for a credence function to be compatible with one's evidence could be defended.

One worry you might have about the revised chance grounding thesis is that in many cases our evidence seems to rule out certain credence assignments as irrational, even though it's difficult to see which chance hypotheses we might appeal to in explaining why this is so. Take for instance the proposition that my friend Jakob will have the extraordinarily spicy *phaal* curry for dinner tonight. I know that he loves spicy food, and I've had *phaal* with him a few times in the past year. In light of my evidence, some credence assignments seem clearly irrational. A value of 0.001 certainly seems too low, and a value of 0.9 certainly seems too high. However, we don't normally think of our credence in propositions of this kind as being constrained by information about chances.

If this is correct, then the revised chance grounding thesis can at best provide a partial account of what it takes for a body of evidence to rule out a credence assignment as irrational. Of course, one could insist that we do have some information about chances which allows us to rule out the relevant credence assignments, but such an idea would have to be worked out in a lot more detail before it could be made plausible. Alternatively, one could simply deny my claim that these credence assignments would be irrational. However, as we'll soon discover, that response would merely strengthen my objection.[9]

Going forward, I will assume that the evidence grounding thesis holds, so that a rational agent's credal state should include all and only those credence functions that are compatible with her total evidence. I will also assume that this notion of compatibility is an objective one, so that there is always a fact of the matter about which credence functions are compatible with a given body of evidence. However, I will not assume any particular understanding of compatibility, such as those provided by White's chance grounding thesis or my revised formulation. As we'll see, these assumptions spell trouble for the

9 Another case where it's not immediately clear how to apply the revised chance grounding thesis is propositions about past events. On what I take to be the standard view, such propositions have an objective chance of either 1 or 0, depending on whether they occurred or not (see for instance Schaffer 2007). So for a proposition A about an event that is known to be in the past, the only chance hypotheses left open by the evidence are (at most) 0 and 1. However, in certain cases, this will be enough to give us maximal imprecision. If we have no knowledge of what the chance of A was prior to the event's occurring (or not occurring), then it seems that any way of distributing credence across these two chance hypotheses will be compatible with our evidence, and hence that the credal state will include a credence function P with $P(A) = x$ for each $x \in [0, 1]$. Indeed, if we accept Levi's (1980, chapter 9) credal convexity requirement, then whenever the credal state includes 0 and 1, it will also include everything in between. A further worry, which I will set aside here, is whether we can have any non-trivial objective chances if determinism is true.

imprecise Bayesian. I will therefore revisit them in Section 6, to see whether they can be given up.

2.4 LOCAL BELIEF INERTIA

In certain cases, evidentially-motivated imprecise Bayesianism makes inductive learning impossible. Joyce already recognizes this, but I will argue that the implications are more wide-ranging and therefore more problematic than has been appreciated so far.[10] To illustrate the phenomenon, consider an example adapted from Joyce (2010:290).

UNKNOWN BIAS A coin of unknown bias is about to be flipped. What is your credence $\mathcal{P}(H_1)$ that the outcome of the first flip will be heads? And after having observed n flips, what is your credence that the coin will come up heads on the $(n+1)$th flip?

As in the Symmetrical Biases example discussed earlier, each $P \in \mathcal{P}$ is here given as the expected value of a corresponding probability density function, f_P, over the possible chance hypotheses. We are not provided with any evidence that bears on the question of whether the first outcome will be heads, and hence our evidence cannot rule out any pdfs as incompatible. In turn, this means that no value of $P(H_1)$ can be ruled out, and therefore that our overall credal state with respect to this proposition will be maximally imprecise: $\mathcal{P}(H_1) = (0,1)$.[11] However, this starting point renders inductive learning impossible, in the following sense. Suppose that you observe the coin being flipped a thousand times, and see 500 heads and 500 tails. This looks like incredibly strong evidence that the coin is very, very close to fair, and would seem to justify concentrating your credence on some fairly narrow interval around 0.5. However, although each element of the credal state will indeed move toward the midpoint, there will always remain elements on each extreme. Indeed, for any finite sequence of outcomes and for any $x \in (0,1)$, there will be a credence function $P \in \mathcal{P}$ which assigns a value of x to the proposition that the next outcome will be heads, conditional on that sequence. Thus your credence that the next outcome will be heads will remain maximally imprecise, no matter how many observations you make.

Bradley (2015) calls this the problem of belief inertia. I will refer to it as local belief inertia, as it pertains to a limited class of beliefs, namely those about the outcomes of future coin flips. This is a troubling implication, but Joyce (2010:291) is willing to accept it:

> if you really know *nothing* about the [...] coin's bias, then you also really know *nothing* about how your opinions about $[H_{n+1}]$ should

10 Joyce is of course not the first to recognize this. See for instance Walley's (1991:93) classic monograph for a discussion of how certain types of imprecise probability have difficulties with inductive learning.

11 Joyce (2010:290) thinks we should understand maximal imprecision here to mean the open set $(0,1)$ rather than the closed set $[0,1]$, but it's not obvious on what basis we might rule out the two extremal probability assignments. At any rate, my objection won't turn on which of these is correct, as we'll see shortly.

> change in light of frequency data. [...] You cannot learn anything in cases of pronounced ignorance simply because a prerequisite for learning is to have prior views about how potential data should alter your beliefs, but you have no determinate views on these matters at all.

Nevertheless, he suggests a potential way out for imprecise Bayesians who don't share his evidentialist commitments. The underlying idea is that we should be allowed to rule out those probability density functions that are especially biased in certain ways. Some pdfs are equal to zero for entire subintervals (a, b), which means that they could never learn that the true chance of heads lies within (a, b). Perhaps we want to rule out all such pdfs, and only consider those that assign a non-zero value to every subinterval (a, b). Similarly, some pdfs will be extremely biased toward chance hypotheses that are very close to one of the endpoints, with the result that the corresponding credence functions will be virtually certain that the outcome will be heads, or virtually certain that the outcome will be tails, all on the basis of no evidence whatsoever. Again, perhaps we want to rule these out, and require that each $P \in \mathcal{P}$ assigns a value to H_1 within some interval (c_-, c^+), with $c_- > 0$ and $c^+ < 1$.

With these two restrictions in place, the spread of our credence is meant to shrink as we make more observations, so that after having seen 500 heads and 500 tails, it is centred rather narrowly around 0.5, thereby making inductive learning possible again. While recognizing this as an available strategy, Joyce does not endorse it himself, as it is contrary to the evidentialist underpinnings of his view. In any case, the strategy doesn't do the trick. Even if we could find a satisfactory motivation, it would not deliver the result Joyce claims it does, as the following theorem shows:

Theorem 1. Let the random variable X be the coin's bias for heads, and let the random variable Y_n be number of heads in the first n flips. For a given n, a given y_n, a given interval (c_-, c^+) with $c_- > 0$ and $c^+ < 1$, and a given $c_0 \in (c_-, c^+)$, there is a pdf, f_X, such that

1. $E[X] \in (c_-, c^+)$,
2. $E[X \mid Y_n = y_n] = c_0$, and
3. $\int_a^b f_X(x)\, dx > 0$ for every $a, b \in [0, 1]$ with $a < b$.

The first and third conditions are the two constraints that Joyce suggested we impose. The first ensures that the pdf is not extremely biased toward chance hypotheses that are very close to one of the endpoints, and the third ensures that it is non-zero for every subinterval (a, b) of the unit interval. The second condition corresponds to the claim that we still don't have inductive learning, in the sense that no matter what sequence of outcomes is observed, for every $c_0 \in (c_-, c^+)$, there will be a pdf whose expectation conditional on that sequence is c_0.

Proof. Consider the class of beta distributions. First, we will pick a distribution from this class whose parameters α and β are such that the first two conditions are satisfied. Now, the expectation and the conditional expectation of a beta distribution are respectively given as

$$E[X] = \frac{\alpha}{\alpha+\beta}, \text{ and } E[X \mid Y_n = y_n] = \frac{\alpha + y_n}{\alpha+\beta+n}.$$

The first two conditions now give us the following constraints on α and β:

$$c_- < \frac{\alpha}{\alpha+\beta} < c^+, \text{ and } \frac{\alpha+y_n}{\alpha+\beta+n} = c_0.$$

The first of these constraints gives us that

$$\frac{c_-}{1-c_-}\beta < \alpha < \frac{c^+}{1-c^+}\beta.$$

The second constraint allows us to express α as

$$\alpha = \frac{c_0(\beta+n) - y_n}{1-c_0}.$$

Putting the two together, we get

$$\beta > \frac{(1-c_-)(y_n - c_0 n)}{c_0 - c_-} \text{ and } \beta > \frac{(1-c^+)(y_n - c_0 n)}{c_0 - c^+}.$$

As we can make β arbitrarily large, it is clear that for any given set of values for n, y_n, c_-, c^+ and c_0, we can find a value for β such that the two inequalities above hold. We have thus found a beta distribution that satisfies the first two conditions. Finally, we show that the third condition is met. The pdf of a beta distribution is given as

$$f_X(x) = \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

where the beta function B is a normalization constant. As is evident from this expression, we will have $f_X(x) > 0$ for each $x \in (0,1)$, which in turn implies that $\int_a^b f_X(x)\,dx > 0$ for every $a, b \in [0,1]$ with $a < b$. Moreover, this holds for any values of the parameters α and β. Therefore every beta distribution satisfies the third condition, and our proof is done. □

What this shows is that all the work is being done by the choice of the initial interval. Although many credence functions will be able to move outside the interval in response to evidence, for every value inside the interval, there will always be a a credence function that takes that value no matter what sequence of outcomes has been observed. Thus the set of prior credence values will be a subset of the set of posterior credence values. The intuitive reason for this is that we can always find an initial probability density function which is sufficiently biased in some particular way to deliver the desired posterior credence value.

There are therefore two separate things going on in the unknown bias case, both of which might be thought worrisome: the problem of maximal imprecision, and the problem of belief inertia. As the result shows, Joyce's proposed fix addresses the former but not the latter, and our beliefs can therefore be inert without being maximally imprecise.[12] Granted, having a set of posterior credence values that always includes the set of prior credence values as a subset is a less severe form of belief inertia than having a set of posterior credence values that is always identical to the set of prior credence values. However, even this weaker form of belief inertia means that no matter how much evidence the agent receives, she cannot converge on the correct answer with any greater precision than is already given in her prior credal state.

Now, Theorem 1 only shows that one particular set of constraints is insufficient to make inductive learning possible in the unknown bias case. Thus some other set of constraints could well be up to the job.

For example, consider the set of beta distributions with parameters α and β such that $\beta/m \leq \alpha \leq m\beta$ for some given number m. If we let the credal state contain one credence function for each of these distributions, inductive learning will be possible.

It may be objected that we should regard belief inertia, made all the more pressing by Theorem 1, not as a problem for imprecise Bayesianism, but rather as a problem for an extreme form of evidentialism. Suppose that a precise Bayesian says that all credences that satisfy the first and third conditions are permissible to adopt as one's precise credences. Theorem 1 would then tell us that it is permissible to change your credence by an arbitrarily small amount in response to any evidence. Although hardcore subjectivists would be happy to accept this conclusion, most others would presumably want to say that this constitutes a failure to respond appropriately to the evidence. Therefore, whatever it is that a precise moderate subjectivist would say to rule out such credence functions as irrational, the imprecise Bayesian could use the same account to explain why those credence functions should not be included in the imprecise credal state.

I agree that belief inertia is not an objection to imprecise Bayesianism as such: it becomes an objection only when that framework is combined with Joyce's brand of evidentialism. Nevertheless, I do believe the problem is worse for imprecise Bayesianism than it is for precise Bayesianism. On the imprecise evidentialist view, you are epistemically required to include all credence functions that are compatible with your evidence in your credal state. If we take Joyce's line and don't impose any further conditions, this means that, in the unknown bias case, you are epistemically required to adopt a credal state that is both maximally imprecise and inert. If we instead are sympathetic to the two further constraints, it means that you are epistemically required to adopt a credal state that will always include the initial interval from which you started as a

12 In turn, this explains why it doesn't matter whether we understand maximal imprecision to mean $(0, 1)$ or $[0, 1]$. Belief inertia will arise regardless of which of the two we choose.

subset. By contrast, on the precise evidentialist view, you are merely epistemically permitted to adopt one such credence function as your own. Of course, we may well think it's epistemically impermissible to adopt such credence functions. But a view on which we are epistemically required to include them in our credal state seems significantly more implausible.

A further difference is that any fixed beta distribution will eventually be pushed toward the correct distribution. Thus any precise credence function will eventually give us the right answer, even though this convergence may be exceedingly slow for some of them. By contrast, Theorem 1 shows that the initial interval (c_-, c^+) will always remain a subset of the imprecise Bayesian's posterior credal state. Therefore, belief inertia would again seem to be more of a problem for the imprecise view than for the precise view.

Finally, it's not at all obvious what principle a precise Bayesian might appeal to in explaining why the credence functions that intuitively strike us as insufficiently responsive to the evidence are indeed irrational. Existing principles provide constraints that are either too weak (for instance the principal principle or the reflection principle) or too strong (for instance the principle of indifference). It may well be possible to formulate an adequate principle, but to my knowledge this has not yet been done.

At any rate, Joyce is willing to accept local belief inertia in the unknown bias case, and his reasons for doing so may strike one as quite plausible. When one's evidence is so extremely impoverished, it might make sense to say that one doesn't even know which hypotheses would be supported by subsequent observations. This case is a fairly contrived toy example, and one might hope that such cases are the exception and not the rule in our everyday epistemic lives. So a natural next step is to ask how common these cases are. If it turns out that they are exceedingly common—as I will argue that they in fact are—then we ought to reject evidentially-motivated imprecise Bayesianism, even if we were initially inclined to accept particular instances of belief inertia.

2.5 GLOBAL BELIEF INERTIA

I will argue that belief inertia is in fact very widespread. My strategy for establishing this conclusion will be to first argue that an imprecise Bayesian who respects the evidence grounding thesis must have a particular prior credal state, and second to show that any agent who starts out with this prior credal state and updates by imprecise conditionalization will have inert beliefs for a wide range of propositions.

As we saw in the previous chapter, in order for the Bayesian machinery—whether precise or imprecise—to get going, we must first have priors in place. In the precise case, priors are given by the credence function an agent adopts before she receives any evidence whatsoever. Similarly, in the imprecise case, priors are given by the set of credence functions an agent adopts as her credal state before she receives any evidence whatsoever. The question of which con-

straints to impose on prior credence functions is a familiar and long-standing topic of dispute within precise Bayesianism. As we have seen, hardcore subjectivists hold that any probabilistic prior credence function is permissible, whereas objectivists wish to narrow down the number of permissible prior credence functions to a single one. In between these two extremes, we find a spectrum of moderate views. These more measured proposals suggest that we add some constraints beyond probabilism, without thereby going all the way to full-blown objectivism.

The same question may of course be asked of imprecise Bayesianism as well. In this context, our concern is with which constraints to impose on the set of prior credence functions. Hardcore subjectivists hold that any set of probabilistic prior credence functions is permissible, whereas objectivists will wish to narrow down the number of permissible sets of prior credence functions to a single one. In between these two extremes, we again find a spectrum of moderate views. For an imprecise Bayesian who is motivated by evidential concerns, the answer to the question of priors should be straightforward. By the evidence grounding thesis, our credal state at a given time should include all and only those credence functions that are compatible with our evidence at that time. In particular, this means that our prior credal state should include all and only those credence functions that are compatible with the empty body of evidence. Thus, in order to determine which prior credal states are permissible, we must determine which credence functions are compatible with the empty body of evidence. As you'll recall, I assumed that the relevant notion of compatibility is an objective one. This means that there will be a unique set of all and only those credence functions that are compatible with the empty body of evidence.[13] Which credence functions are these?

In light of our earlier examples, we can rule out some credence functions from the prior credal state. In particular, we can rule out those that don't satisfy the principal principle. If we were to learn only that the chance of A is x, then any credence function that does not assign a value of x to A will be incompatible with our evidence. And given that the credal state is updated by conditionalizing each of its elements on all of the evidence received, it follows that we must have $P(A|\text{ch}(A) = x) = x$ for each P in the prior credal state $\mathcal{P}_0$. Along these lines, some may also wish to add other deference principles.

Now, one way of coming to know the objective chance of some event seems to be via inference from observed physical symmetries.[14] If that's right, it would appear to give us a further type of constraint on credence functions in the prior credal state. More specifically, if some proposition *Symm* about physical symmetries entails that $\text{ch}(A) = x$, then all credence functions P in the prior credal state should be such that $P(\text{ch}(A) = x \mid \mathit{Symm}) = 1$. Given that we've accepted the principal principle, this means that we also get that $P(A \mid \mathit{Symm}) = x$. Now,

13 This objectivism may strike you as implausible or undesirable. In the next section, we will consider whether an imprecise Bayesian can give it up without also giving up their evidentialist commitment.

14 I'm grateful to Pablo Zendejas Medina for emphasizing this.

what sort of things do we have to include in Symm in order for the inference to be correct? In the case of a coin flip, we presumably have to include things like the coin's having homogenous density together with facts about the manner in which it is flipped.[15] But given that we are trying to give *a priori* constraints on credence functions, it seems that this cannot be sufficient. We must also know that, say, the size of the coin or the time of the day are irrelevant to the chance of heads, and similarly for a wide range of other factors. Far-fetched as these possibilities may be, it nevertheless seems that we cannot rule them out *a priori*.

I will return to a discussion of the role of physical symmetries shortly. For the moment, it suffices to note that symmetry considerations, just like the principal principle and other deference principles, can only constrain conditional prior credence assignments, leaving the whole range of unconditional prior credence assignments open. Are there any legitimate constraints on unconditional prior credence assignments? As we have seen, some endorse the regularity principle, which requires credence functions to assign credence 0 only to propositions that are in some sense (usually doxastically) impossible. So perhaps we should demand that all credence functions in the prior credal state be regular.

So far, I've surveyed a few familiar constraints on credence functions. The thought is that if we add enough of these, we may be able to avoid many instances of belief inertia. However, this strategy faces a dilemma: on the one hand, adding more constraints means that we are more likely to successfully solve the problem. On the other, the more constraints we add, the more it looks like we're going beyond our evidence, in much the same way that the principle of indifference would have us do. Given that Joyce endorsed imprecise Bayesianism for the very reason that it allowed us to avoid having to go beyond the evidence in this manner, this would be especially problematic. Let us therefore assume that the only constraints we can impose on the credence functions in our prior credal state are the principal principle and other deference principles, constraints given by symmetry considerations, and possibly also the regularity principle. This gives us the following result. The evidence grounding thesis, together with an objective understanding of compatibility, imply:

MAXIMALLY IMPRECISE PRIORS For any contingent A, a rational agent's prior credence $\mathcal{P}_0(A)$ in that proposition is maximally imprecise.[16]

Why does this follow? Take an arbitrary contingent proposition A. If we accept the regularity principle, the extremal credence assignments 0 and 1 are of course ruled out. The principal principle and other deference principles only constrain conditional credence assignments. For example, the principal principle requires each P in the prior credal state $\mathcal{P}_0$ to satisfy $P(A \mid \mathrm{ch}(A) = x) = x$, where $\mathrm{ch}(A) = x$ is the proposition that the objective chance of A is x. Other deference principles have the same form, with $\mathrm{ch}(\cdot)$ replaced by some

15 See Strevens (1998) for one account of how this works in more detail.

16 Where 'maximally imprecise' means either $\mathcal{P}_0(A) = (0,1)$ or $\mathcal{P}_0(A) = [0,1]$, depending on whether or not we accept the regularity principle.

other probability function one should defer to. By the law of total probability for continuous variables, we have that

$$P(A) = \int_0^1 P(A \mid \mathrm{ch}(A) = x) \cdot f_P(x)\, dx,$$

where $f_P(x)$ is the pdf over possible chance hypotheses that is associated with P. By the principal principle, it follows for all values of x that $P(A \mid \mathrm{ch}(A) = x) = x$, which in turn means that

$$P(A) = \int_{-\infty}^{\infty} x f_P(x)\, dx.$$

This means that the value of $P(A)$ is effectively determined by the pdf $f_P(x)$. Therefore, if we are to use the principal principle to rule out some assignments of unconditional credence in A, we have to do so by ruling out, *a priori*, some pdfs over chance hypotheses. Given the constraints we have accepted on the prior credal state, the only way of doing this[17] would be via symmetry considerations. However, in order to do so we would first have to rule out certain credence assignments over the various possible symmetry propositions. As we have no means of doing so, it follows that neither the principal principle nor symmetry considerations allow us to rule out any values for $P(A)$. Any other deference principles will have the same formal structure as the principal principle, and the corresponding conclusions therefore hold for them as well. We thus get maximally imprecise priors.

Next, we will examine how an agent with maximally imprecise priors might reduce their imprecision. Before doing that, however, I'd like to address a worry you might have about the inference to Maximally Imprecise Priors above. I have been speaking of prior credal states as if they were just like posterior credal states, the only difference being that they're not based on any evidence. But of course, the notion of a prior credal state is a fiction: there is no point in time at which an actual agent adopts it as her state of belief. And given that my formulation of the evidence grounding thesis makes it clear that it is meant to govern credal states at particular points in time, we have no reason to think that it also applies to prior credal states.

If the prior credal state is a fiction, what kind of a fiction is it? Titelbaum (manuscript, p. 110) suggests that we think of priors as encoding an agent's ultimate evidential standards.[18] Her ultimate evidential standards determine how she interprets the information she receives. In the precise case, an agent whose credence function at t_1 is P_1 will regard a piece of evidence E_i as favouring a proposition A if and only if $P_1(A|E_i) > P_1(A)$. So her credence function P_1 gives us her evidential standards at t_1. Of course, her evidential standards in this sense will change over time as she obtains more information. It may be that in between t_1 and t_2 she receives a piece of evidence E_2 such that

17 Other than the uninteresting case of the regularity principle ruling out discontinuous pdfs that concentrate everything on the endpoints 0 and 1.

18 This kind of view of priors is of course not original to Titelbaum. See for example Lewis (1980:288).

$P_2(A|E_i) < P_2(A)$. If she does, at t_2 she will no longer regard E_i as favouring A. In order to say something about how she is disposed to evaluate total bodies of evidence, we must turn to her prior credence function, which encodes her ultimate evidential standards. If an agent with prior credence function P_0 has total evidence E, she will again regard that evidence as favouring A if and only if $P_0(A|E) > P_0(A)$. In the same way, we can think of a prior credal state as encoding the ultimate evidential standards of an imprecise agent.[19]

Suppose that we have a sequence of credence functions $P_1, P_2, P_3, \ldots$, where each element P_i is generated by conditionalizing the preceding element P_{i-1} on all of the evidence obtained between t_{i-1} and t_i. We will then be able to find a prior credence function P_0 such that, for each P_i in the sequence, $P_i(\cdot) = P_0(\cdot|E_i)$, where E_i is the agent's total evidence at t_i. Because a credal state is just a set of credence functions, we will also be able to find a prior credal state $\mathcal{P}_0$ such that the preceding claim holds of each of its elements.[20]

This means that, in order to arrive at Joyce's judgements about particular cases, we must make assumptions about the prior credal state as well. Consider for instance the third urn example, where we don't even know what colours the marbles might have. If we are to be able to say that it is irrational to have a precise credence in B_3 (the proposition that a marble drawn at random from this urn will be black), we must also say that it is irrational to have a prior credal state $\mathcal{P}_0$ such that there is an x such that $P(B_3|E) = x$ for each $P \in \mathcal{P}_0$, where E is the (limited) evidence available to us (namely that the urn contains one hundred marbles of unknown colours, and that one will be drawn at random). Similarly, in the unknown bias case, we must rule out as irrational any prior credal state which does not yield the verdict of maximal imprecision.

So although the prior credal state is in a certain sense fictitious, the evidence grounding thesis must still apply to it, if it is to apply to posterior credal states at all. Because of the intimate connection (via imprecise conditionalization on the total evidence) between the prior credal state and posterior credal states, any claims about the latter will imply claims about the former. Therefore, if the evidence grounding thesis is to constrain an agent's posterior credal states, it must also constrain her ultimate evidential standards, namely her prior credal state. Thus the argument for maximally imprecise priors still stands.

In order to determine how widespread belief inertia is, we must now consider how an agent with maximally imprecise priors might reduce her imprecision

19 In this case, we will have to say a bit more about what it means for an agent to regard a piece of evidence as favouring a proposition. Presumably a supervaluationist account, along the lines of the one we sketched for unconditional comparative judgements, will do: an agent with credal state $\mathcal{P}$ will regard a piece of evidence E_i as determinately favouring A if and only if $P(A|E_i) > P(A)$ for each $P \in \mathcal{P}$.

20 Now, P_i and E_i will not determine a unique P_0. There will be distinct P_0 and P_0' such that $P_i(\cdot) = P_0(\cdot \mid E_i)$ and $P_i(\cdot) = P_0'(\cdot \mid E_i)$. In the case of an imprecise Bayesian agent, this means that we cannot infer her prior credal state from her current credal state together with her current total body of evidence. However, given that we are for the moment assuming that the notion of compatibility is an objective one, the prior credal state $\mathcal{P}_0$ should consist of all and only those credence functions that satisfy the relevant set of constraints, and hence that $\mathcal{P}_0$ will be unique.

with respect to some particular proposition. One obvious way for her to do so is through learning the truth of that proposition. If she learns that A, then all credence functions in her posterior credal state will agree that $P(A) = 1$. Given that we required all credence functions in the prior credal state to satisfy the principal principle, another way for the agent to reduce her imprecision with respect to A is to learn something about the chance of A. If she learns that $\mathrm{ch}(A) = x$, then all credence functions in her posterior credal state will agree that $P(A) = x$. Similarly, if she learns that the chance of A lies within some interval $[a, b]$, then all of them will assign a value to A that lies somewhere in that interval.[21] And if we take other deference principles on board as well, those will yield analogous cases.

Although knowledge of objective chance is a staple of probability toy examples, how often do we come by such knowledge in real life? The question is all the more pressing for the imprecise Bayesian. As the unknown bias case illustrated, if an imprecise Bayesian starts out with no information about the objective chance of some class of events, she cannot use observed outcomes of events in this class to narrow down her credence. By contrast, precise Bayesians can use such information to obtain a posterior credence that will eventually be within an epsilon of the objective chance value.

As discussed earlier, we do have one other way of obtaining information about objective chance, namely via inference from physical symmetries. Now, the question is: how often are we in a position to conditionalize on propositions about such symmetries? First, and most obviously, the principle will only be able to constrain credences in propositions for which the relevant physical symmetries are present. Thus even if we are happy to say that the proposition that my friend Jakob will have *phaal* curry for dinner tonight, or the proposition that the next raven to be observed will be black have non-trivial objective chances, there are presumably no physical symmetries to rely on here. Hence the principle has limited applicability.

Second, in cases where the relevant physical symmetries do exist, we must also know that other factors are irrelevant to the objective chance, as mentioned earlier. From our everyday interactions with the world, as well as from physical theory, we know that the size of a coin and the time of the day are irrelevant to the chance of heads. But how might our imprecise Bayesian accommodate this datum? We know from before that she will have a maximally imprecise prior in any contingent proposition, and hence in any physical theory. So in order to make use of these physical symmetries, she must first narrow down the range of these credences, and assign higher credence to theories according to which the irrelevant factors are indeed irrelevant.

21 I have not explained how the update works when an agent learns that the chance of A lies within some interval $[a, b]$. One way of doing this is to set each pdf f_P to equal zero everywhere outside of that interval and then normalize it, so that $\int_a^b f_P(x)\,dx = 1$. Although I don't believe much of my argument turns on it, there are other ways of doing this as well.

But this brings us back to the same problem: how can the imprecise Bayesian reduce her imprecision with respect to these physical theories? Even if we think it's intelligible to think of physical theories as having objective chance of being true, it seems clear that we'll never be in a position to conditionalize on propositions about their objective chance. Furthermore, given that physical theories make claims that go beyond one's evidence, we cannot directly conditionalize a physical theory itself. Thus it would appear that, in practice, the imprecise Bayesian cannot use symmetry considerations to reduce her imprecision. I take it as a given that we do have some way of rationally narrowing down the range of possible objective chance values. We may not know their exact values, but we can nevertheless do a lot better than forever remaining maximally imprecise. The challenge for the evidentially-motivated imprecise Bayesian is to explain how this is possible within their framework.

As you will recall, I suggested that we might want to take on board deference principles other than the principal principle. So a further way of reducing one's imprecision with respect to some proposition would be to defer to a relevant expert. To do so, we must say a bit more about who counts as an expert. The first thing to note here is that if someone has arrived at a relatively precise credence in A through reasoning that is not justified by the lights of evidentially-motivated imprecise Bayesianism, she cannot plausibly count as an expert with respect to A. If the precision of her credence goes beyond her evidence in an unwarranted way, the same must hold of anyone who defers to her credence as well. This greatly limits the applicability of the deference principle. Therefore, we can only legitimately defer to experts in cases where those experts have conditionalized on A directly.[22] However, in order to do so we must not only know what the expert's credence in A is, but also that she is indeed an expert. And again, we don't seem to have a way of narrowing down our initial, maximally imprecise credence that this person is an expert with respect to A.

Given that the constraints we accepted on prior conditional credence assignments have such limited practical applicability, we get the following result:

GLOBAL BELIEF INERTIA For any proposition A, a rational agent will have a maximally imprecise credence in A unless her evidence logically entails either A or its negation.

Even if we were willing to concede some instances of local belief inertia, such as in the unknown bias case, this conclusion should strike us as unacceptable. It invalidates a wide range of canonically rational comparative confidence judgments. Propositions that are known to be true are assigned a credence of 1, those that are known to be false are assigned a credence of 0, and all others are assigned a maximally imprecise credence. Although some comparative confidence judgments will remain intact—for instance, all credence functions

22 As well as in cases where the expert herself bases her credence on that of another expert, along a sequence of deferrals that must eventually end with someone who conditionalized on A directly.

will regard four heads in a row as more likely than five heads in a row—many others will not. Surely a theory of inductive inference should do better.[23]

2.6 RESPONSES

In a sense, global belief inertia is hardly a surprising result in light of my strong assumptions. I assumed the evidence grounding thesis, which states that the credal state must contain all and only those credence functions that are compatible with the evidence. Moreover, I assumed that compatibility is an objective notion, so that there is always an agent-independent fact of the matter as to whether a particular credence function is compatible with a given body of evidence. Finally, I noted that compatibility must be very permissive (in the sense of typically counting a wide range of credence functions as compatible with any particular body of evidence), because otherwise we risk making the same mistake as the one we accused the principle of indifference of making. With all of these assumptions on board, it's almost a given that global belief inertia follows. The question is whether we can motivate imprecise Bayesianism on the grounds that precise credences are often epistemically reckless because they force us to go beyond our evidence, without having the resulting view fall prey to global belief inertia.

Some technical fixes may solve the problem. We saw that Joyce's suggestion for how to avoid belief inertia in the unknown bias case didn't do the job, but perhaps an approach along similar lines could be made to work.[24] However, as Joyce concedes, such a proposal could not be justified in light of his evidentialist commitments. Similarly, we might try replacing imprecise conditionalization with some other update rule that allows us to move from maximal imprecision to some more precise credal state. One natural idea is to introduce a threshold, so that credence functions which assigned a value below that threshold to a proposition that we then go on to learn, get discarded from the posterior credal state: $\mathcal{P}_1 = \{P(\cdot \mid E_1) : P \in \mathcal{P}_0 \wedge P(E_1) > t\}$.[25] The threshold proposal comes with problems of its own: it violates the commutativity of evidence (the order in which we learn two pieces of evidence can make a difference for which credal state we end up with), and it may lead to cases where the credal state becomes the empty set. But again, the more fundamental problem is that it violates the evidentialist commitment. By discarding credence functions that don't meet the threshold, we go beyond the evidence.

In general, the dilemma for evidentially-motivated imprecise Bayesianism is that in order to avoid widespread belief inertia, we must either place stronger constraints on the uniquely rational prior credal state, or concede that there is a range of different permissible prior credal states. However, these two

23 See Rinard (2013) for further discussion of the implications of maximal imprecision for comparative confidence judgments.

24 I mentioned one such idea in the context of the unknown bias case: let all the credence functions be based on beta distributions whose parameters are restricted in a particular way.

25 This threshold rule is mentioned by Bradley and Steele (2014). A related method is the maximum likelihood rule given by Gilboa and Schmeidler (1993).

strategies expose the view to the same criticism that we made of objective and subjective precise Bayesianism: they allow agents to go beyond their evidence.

You might worry that the argument for global belief inertia relied on a tacit assumption that the only way of spelling out the underlying evidentialism is via some connection to objective chance (as done, for example, by the chance grounding theses). Once we see that this leads to Global Belief Inertia, we should give up that view, but that doesn't mean we have to give up the evidentialism itself. Indeed, even in the absence of a detailed account of how evidence constrains credal states, it seems quite obvious that our current evidence does not support a precise credence in, say, the proposition that there will be four millimeters of precipitation in Paris on 3 April 2237. So the case for evidentially-motivated imprecision still stands.

The claim is not merely that there is no unique precise credence that is best supported by the evidence. If it were, precise Bayesians could simply respond by saying that there are multiple precise credences, each of which one could rationally adopt in light of the evidence. Instead, the claim must be that, on its own, any precise credence would be an unjustified response to the evidence. Hence the evidence only supports imprecise credences. But does it support a unique imprecise credence, or are there multiple permissible imprecise credences? On the face of it, the claim that it supports a unique imprecise credence looks quite implausible. At any rate, it is a claim that stands in need of further motivation. The revised chance grounding thesis gave us one possible explanation of this uniqueness. By including credence functions in the credal state on the basis of their consistency with what we know about objective chance, our criterion gives a clear-cut answer in every case, and hence uniqueness follows. But now that we've rejected the revised chance grounding thesis because of the widespread belief inertia it gave rise to, we no longer have any reason to suppose that the evidence will always support a unique credal state. In the absence of a more detailed account of evidential support for credal states, we should reject uniqueness.

Suppose therefore that we instead accept that our evidence supports multiple imprecise credences. On what grounds can we then say that it doesn't also support some precise credences? The intuition behind the thought that no precise credence is supported by the evidence also suggests that, for sufficiently small values of ϵ, no imprecise credence of $[x - \epsilon, x + \epsilon]$ is supported by the evidence, so the relevant distinction cannot merely be between precise and imprecise credences. What the intuition suggests is instead presumably that no credence that is too precise is supported by the evidence, whether this be perfect precision or only something close to it. But again, to say what qualifies as too precise, we need a more detailed account of evidential support for credal states.

At this point, my interlocutor might simply reiterate their original point, cast in a slightly new form. Yes, they will say, we don't know exactly which credences are too precise for our evidence. But even though we don't have a detailed

account, it is still quite clear that some credences are too precise whereas others aren't. So the case for evidentially-motivated imprecision still stands. To give this idea a bit more flesh, consider an analogy with precise Bayesianism. Unless they are thoroughly subjectivist, precise Bayesians hold that some prior credence functions are rational and others aren't. For example, stubborn priors that are moved an arbitrarily small amount even by large bodies of evidence may well be irrational. This cannot be explained by any evidence about objective chance, or indeed by any other kind of evidence, because by definition priors aren't based on any evidence. There are just facts about which of them are rational and which aren't. Furthermore, a credence function is supported by a body of evidence just in case it is the result of conditionalizing a rational prior on that body of evidence.[26] Now, imprecise Bayesians can say the same of their view. Some imprecise prior credal states are rational and others aren't. Again, this cannot be based on any evidence about objective chance, because prior credal states aren't based on any evidence. There are just facts about which of them are rational and which aren't. Furthermore, a credal state is supported by a body of evidence just in case it is the result of conditionalizing a rational prior credal state on that body of evidence.

I won't attempt to resolve this large dispute here, so let me just say two things in response. The first is simply that those who follow Joyce's line of argument is unlikely to be happy with this kind of position, given that it appears to be vulnerable to the same criticisms as those he raised for precise objective Bayesianism. Of course, imprecise Bayesians who don't share these commitments may well want to respond along these lines, which brings me to my second point: even if they can't give us an exact characterization of which imprecise priors are permissible, they should at least be able to show that none of the permissible priors give rise to widespread belief inertia. Before that has been done, it seems premature to think that the problem has been solved.

Before concluding, let me briefly explore some other tentative suggestions for where to go from here. If we wish to keep the formal framework as it is (namely imprecise probabilism and imprecise conditionalization, together with the supervaluationist understanding of credal states), then one option is to scale back our ambitions. Instead of saying that imprecise credences are rationally required in, say, the second and third urn cases, we only say that they are among the permissible options.

This response constitutes a significant step in the direction of subjectivism. We can still place some constraints on the credence functions in the prior credal state (for example that they satisfy the principal principle). But instead of requiring that the prior credal state includes all and only those credence functions that satisfy the relevant constraints, we merely require that it includes only (but not necessarily all) credence functions that satisfy them. On this view, precise Bayesianism goes wrong not in that it forces us to go beyond our evidence (any view that avoids belief inertia will have to!), but rather because

26 See Williamson (2000, chapter 10) for an example of a view of this kind, cast in terms of evidential probability.

it forces us to go far beyond our evidence, when other more modest leaps are also available. How firm conclusions we want to draw from limited evidence is in part a matter of epistemic taste: some people will prefer to go out on a limb and assign relatively precise credences, whereas others are more cautious, and prefer to remain more non-committal. Both of these preferences are permissible, and we should therefore give agents some freedom in choosing their level of precision.

Another option is to enrich the formal framework in a way that provides us with novel resources for dealing with belief inertia. For example, we might associate a weight with each credence function in the credal state and let the weight represent the credence functions degree of support in the evidence.[27] By letting the weights change in response to incoming evidence, inductive learning becomes possible again, even in cases where the spread of values assigned to a proposition by elements of the credal state remains unchanged. In a similar vein, Bradley (2017) suggests that we introduce a confidence relation over the set of an agent's probability judgements.[28] For example, after having observed 500 heads and 500 tails in the unknown bias case, we may be more confident in the judgement that the probability of heads is in $[0.48, 0.52]$ than we are in the judgement that it is in $[0.6, 1]$. Needless to say, the details of these proposals have to be worked out in much greater detail before we can assess them. Nevertheless, they look like promising options for imprecise Bayesians to explore in the future.

2.7 CONCLUSION

I have argued that evidentially motivated imprecise Bayesianism entails that, for any proposition, one's credence in that proposition must be maximally imprecise, unless one's evidence logically entails either that proposition or its negation. This means that the problem of belief inertia is not confined to a particular class of cases, but is instead completely general. I claimed that even if one is willing to accept certain instances of belief inertia, one should nevertheless reject any view which has this implication. After briefly looking at some responses, I tentatively suggested that the most promising options are either (i) to give up objectivism and concede that the choice of a prior credal state is largely subjective, or (ii) to enrich the formal framework with more structure.

27 See Gärdenfors and Sahlin (1982) for an approach along these lines.
28 This approach is inspired by Hill (2013).

3

MORAL UNCERTAINTY AND ARGUMENTS FOR PROBABILISM

3.1 INTRODUCTION

Our next variation concerns the objects of credence. Typically, Bayesians have been concerned with uncertainty regarding descriptive states of affairs. But it seems we may also be *morally* uncertain: that is, uncertain regarding the moral states of affairs. For example, I may be uncertain whether abortion is morally permissible, whether it's better to order vegetarian, or whether modesty is a virtue. And I may be uncertain about these things even if I am certain of all relevant empirical facts, e.g. facts concerning the cognitive development of the fetus, the conditions of factory-farmed animals, or how a modest person behaves. If so, my uncertainty is fundamentally moral: I'm uncertain of what moral reasons the descriptive facts known to me give rise to.

On the face of it, it seems plausible both that we are in fact sometimes morally uncertain, and that it is sometimes rationally permissible to be morally uncertain. To be sure, both of these claims can be resisted. But I take them to be compelling enough to form a natural starting point for our investigations. And once we allow for the notion of moral uncertainty, we can ask various epistemological questions about it. The main focus of this chapter is one such question: does probabilism hold for our degrees of belief in moral claims?

We shall approach this question by investigating whether the three types of arguments that many take to establish probabilism with respect to ordinary, descriptive uncertainty can also support this doctrine with respect to moral uncertainty. Representation theorem arguments impose rationality constraints on relational attitudes (either preferences or comparative confidence judgments) and show that if the constraints in question are satisfied, the relation can be represented using a probability function. Dutch book arguments show that, given certain assumptions about betting behaviour, agents with non-probabilistic credences are willing to accept a set of bets that together guarantee a sure loss for them. And accuracy arguments show that, given certain assumptions about how to measure distance from truth, probabilistic credences are guaranteed to be closer to truth than non-probabilistic credences.

However, before we can evaluate these arguments, we must say a bit more about the nature of moral uncertainty. In particular, we must get clear on what the objects of credence are supposed to be in the case of moral uncertainty.

The next section provides general background on moral uncertainty. In section 3.3, I propose a way of representing moral claims formally so as to make them suitable as objects of credence. Sections 3.4–3.6 examine the prospects for representation theorem arguments, Dutch book arguments, and accuracy arguments for probabilism with respect to moral uncertainty.

3.2 MORAL UNCERTAINTY

Existing discussions of moral uncertainty have tended to focus on its practical dimension, i.e. on the question of how to make decisions in light of one's moral uncertainty. Although my concerns are primarily epistemological, the two projects are not unrelated, and it will therefore be useful to first go through some of the general issues to do with moral uncertainty.[1]

Consider the following example of decision-making under moral uncertainty. You are deciding whether or not to have meat for dinner, and your credence is divided between speciesism according to which animal welfare doesn't matter, and non-speciesism, according to which animal welfare does matter. The values assigned to the different options by the two theories are as follows:

	Meat	No meat
Speciesism	15	10
Non-Speciesism	-100	10

Furthermore, suppose that your credence in speciesism is 0.9 and that your credence in non-speciesism is 0.1. What should you do in light of this uncertainty? According to one natural line of thought, given that you have much greater confidence in speciesism than in non-speciesism, you should act upon the former and eat meat.[2] However, this mode of reasoning is insensitive the fact that the stakes are much higher for non-speciesism. One way of accommodating this observation is to move to an expected value framework, where the expected value of an option is given as the sum of the values assigned to that option by the relevant theories, weighted by the agent's credences in those theories. In our example, the expected value of eating meat is $15 \cdot 0.9 - 100 \cdot 0.1 = 3.5$, whereas the expected value of not eating meat is 10. So even though you are much more confident in speciesism, expected value reasoning suggests that you ought nevertheless to refrain from eating meat, because otherwise you run the risk of doing something gravely morally wrong.[3]

1 See Hudson (1989), Gracely (1996), Lockhart (2000), Ross (2006), Sepielli (2009, 2010, 2013), Guerrero (2007), Moller (2011), MacAskill (2014), and Bykvist (2017).

2 The idea that one should act in accordance with whatever theory one has highest credence in is sometimes known as *My Favourite Theory*. See Gustafsson and Torpman (2014) for a defense.

3 Weatherson (2014) argues against this kind of moral hedging, and Harman (2015) argues against the general view that moral uncertainty is relevant to what one ought to do.

3.2.1 *Problems for Moral Uncertainty*

As illustrated by this example, the project of developing an account of decision-making under moral uncertainty faces a number of obstacles. First, if non-cognitivism about the moral domain is correct, it's not obvious whether the very idea of moral uncertainty is even intelligible. If moral claims are not truth-apt, it's not clear what it would mean to say that you are uncertain about whether it's permissible to eat meat. Thus the whole project may not even get off the ground in the first place.[4]

Second, if we regard a theory of decision-making under moral uncertainty as an account of what agents ought to do in light of this uncertainty, what notion of 'ought' are we employing? In one sense, what we ought to do under moral uncertainty is simply what the correct moral theory tells us to do. But of course, this is not especially helpful or action-guiding for an agent who is highly uncertain. So can we develop a suitable notion of 'ought' which is sensitive to our moral uncertainty?[5]

Third, some accounts of decision-making under moral uncertainty require us to make intertheoretic comparisons of value; comparisons between how valuable an action is according to one moral theory and how valuable it is according to another moral theory. For example, in calculating expected value of eating meat, I assumed that we can make intertheoretic comparisons between speciesism and non-speciesism. But on the face of it, such comparisons seem arbitrary. How can we compare the value of saving a life on a utilitarian theory with the disvalue of telling a lie on a deontological theory? The problem appears to arise even in the case of theories that are overwhelmingly similar, such as a utilitarian theory and a prioritarian theory. How does the utilitarian value of increasing the well-being of someone well off by some amount compare to the prioritarian value of increasing the well-being of someone badly off by the same amount? The theories themselves do not seem to come equipped with answers to such questions.[6]

Fourth, if we can be uncertain over first-order moral theories, we can presumably be uncertain over second-order moral theories as well, i.e. over theories of decision-making under first-order moral uncertainty. How should we act in light of this second-order moral uncertainty? Well, to answer that we need a third-order moral theory, and an infinite regress appears to arise.[7]

---

4 Smith (2002) and Bykvist and Olson (2009, 2012) argue that non-cognitivists have trouble accounting for moral uncertainty, whereas Sepielli (2011) argues that they can rise to the challenge.

5 See Sepielli (2012) for one account.

6 For some proposals of how to make intertheoretic comparisons, see Lockhart (2000:84), Sepielli (2009: 12), and MacAskill (2014, chapter 4). Gustafsson and Torpman argues for *My Favourite Theory* largely on the basis of (what they take to be) the impossibility of intertheoretic comparisons. MacAskill (2016b) presents an account of decision-making under moral uncertainty that draws on social choice theory rather than expected utility theory, thereby circumventing the need for intertheoretic comparisons.

7 See Sepielli (2013) for discussion.

Given that our concern is with the epistemic rather than practical aspect of moral uncertainty, we can mostly set aside these questions. We are not providing a theory of decision-making under moral uncertainty, and hence the question of which notion of "ought" we are appealing to will not arise.[8] Although the question of intertheoretic comparisons will for the most part not concern us, it does play a central role in the discussion of decision-theoretic representation theorem arguments, so we shall return to the matter in section 3.4.1. We shall not concern ourselves with the regress problem, but the question of non-cognitivism clearly does have bearing on the epistemic aspect of moral uncertainty, so let us consider it in a bit more detail.

3.2.2 *Moral Uncertainty and Non-Cognitivism*

According to moral non-cognitivism, moral claims are not truth-apt. When people make moral statements, they are not expressing their beliefs but rather expressing some non-cognitive attitude. If non-cognitivism is correct, it may seem strange to speak of moral uncertainty. Being uncertain about some claim is usually understood as being uncertain about its truth value. But if moral claims lack truth value, what could it mean to be morally uncertain?

On the simplest expressivist analysis, agents who utter moral sentences do so to express their approval or disapproval. For example, an agent who utters the sentence "Stealing is wrong" thereby expresses their disapproval of stealing. But Smith (2002) argues that no expressivist analysis can simultaneously account for (i) the importance we take a moral judgment to have, (ii) how confident we are in that judgment, and (iii) the stability with which we hold that judgment over time. With regards to importance, I may judge that murder and theft are both morally wrong but hold that murder is much worse than theft. With regards to confidence, I may be much more confident that murder is morally wrong than that abortion is, while believing that if they are both morally wrong, then they are equally wrong. And finally, with regards to stability, I may be equally confident of two acts that they are morally wrong and believe that if they are both morally wrong then they are equally wrong, and yet the former judgment may be much more stable than the latter, in the sense that incoming information is much more likely to make me revise the latter judgment.

How would a cognitivist account for these features? First, the judgment that murder and theft are both morally wrong, but that murder is much worse than theft can be spelled out in terms of a moral theory (or perhaps a set of moral theories) which gives this verdict. Smith's second and third points are, of course, very naturally captured in a Bayesian framework. The confidence

8 It is of course true that the theory we shall consider, i.e. probabilism with respect to moral uncertainty, deals in oughts in the sense that it is a normative epistemological theory. However, the ought in "Your degrees of belief in moral claims ought to be probabilistically consistent" is presumably of the same sort as the ought in "Your degrees of belief in descriptive claims ought to be probabilistically consistent," and there is therefore no new notion in need of explanation here.

with which I make a moral judgment would then correspond to my probabilistic credence in that judgment. And the stability with which I hold a moral judgment could then be spelled out as follows: for some set $\{E_i\}$ of propositions I might learn, my credence in A is more stable than my credence in B just in case, for each E_i, $\mid P(A)P(A \mid E_i) \mid < \mid P(B)P(B \mid E_i) \mid$ (cf. Leitgeb's (2017) stability theory of belief.)

For a simple expressivist, however, things are not so easy. To capture importance, we might say that I disapprove much more strongly of murder than of theft. But how would we then capture confidence? We could try to say that if I am more confident in one moral judgment than another, then I am more likely to hold on to that judgment over time. But of course, that corresponds more closely to stability than to confidence. The basic issue is that in order to capture all three aspects that Smith identifies, we need there to be two gradable features of moral judgments, but the simple expressivist framework only provides us with one.

Of course, this difficulty with moral uncertainty is not the only problem for expressivism. One of its main challenges is explaining how it is that moral sentences behave much like propositions in many ways. For example, one moral sentence may entail another, two moral sentences may be inconsistent, and we can conjoin moral sentences with descriptive propositions. This is the so-called Frege-Geach problem. Let us consider a version of it involving negation (Schroeder 2008):

1. I think that stealing is wrong.
2. I don't think that stealing is wrong.
3. I think that stealing is not wrong.
4. I think that not stealing is wrong.

Here is how a simple expressivist would translate these:

E1. I disapprove of stealing.

E2. I don't disapprove of stealing.

E3. ?

E4. I disapprove of not stealing.

The trouble is that there are three places in sentence (1) where one can insert a negation, but only two places in (E1) where one can do so. As a result, the simple expressivist lacks an analysis of sentence (3). However, more sophisticated expressivists purport to provide adequate analyses. I do not have the space here for an extended discussion of the various expressivist accounts, but by way of illustration I will consider just one example.

Schroeder (2008) introduces the attitude of "being for" as the general conative attitude to figure in expressivist analyses. A moral statement A expresses

For(a), where a is an appropriate analysis of what A says that the speaker is for. So $\neg A$ expresses For($\neg a$), $A \wedge B$ expresses For($a \wedge b$), and so forth. We can then translate (1)-(4) above as follows:

S1. I'm for blaming for stealing.

S2. I'm not for blaming for stealing.

S3. I'm for not blaming for stealing.

S4. I'm for blaming for not stealing.

We also need an account of how to combine moral claims with descriptive claims. Schroeder proposes that descriptive claims can also be analysed in terms of being for: belief in the descriptive claim A is analysed as "being for proceeding as if A." Of course, whether Schroeder's analysis succeeds will depend on whether every use of a moral statement A can plausibly be analysed as expressing that the speaker is for some a. And it also depends on what exactly being for amounts, and why it is inconsistent to both be for blaming for stealing and be for not blaming for stealing. But suppose for the moment that we these issues can be dealt with.

Sepielli (2011) claims that any expressivist theory adequate to handle the Frege-Geach problem is adequate to handle moral uncertainty. However, few of those who argue that expressivism and moral uncertainty are compatible then go on to examine which form this uncertainty will take, and in particular whether something like probabilism can be justified as a requirement of rationality given an expressivist understanding of moral uncertainty.

A recent exception is Staffel (forthcoming), who proposes the following two-step procedure for including moral uncertainty in an expressivist framework. First: find some feature of the framework that admits of degrees. Second: formulate and defend rationality requirements for this feature. In Schroeder's case, Staffel proposes that we introduce the notion of *degrees* of being for. Suppose that I have 0.5 credence that a coin will land heads and 0.5 credence that it will land tails. On Schroeder's account, this means that I'm for proceeding as if the coin will land heads to degree 0.5, and I'm for proceeding as if it will land tails to degree 0.5. Similarly, if I have 0.5 credence that stealing is wrong, then on Schroeder's account I am for blaming for stealing to degree 0.5. Clearly, more needs to be said about what it means for an agent to be for something to degree x. But I shall not do so here.

The arguments for probabilism I will consider in this chapter make most intuitive sense on a cognitivist understanding of moral discourse, although it may be possible to give them non-cognitivist readings. Although I shall sometimes comment on the possibility of giving a non-cognitivist interpretation of some notions that figure in the arguments, I will not provide a systematic account of moral uncertainty for non-cognitivists. Any conclusions are therefore best

read as holding conditional on cognitivism.[9]

This concludes our initial survey of moral uncertainty. Let me make one final clarification. If we are prepared to take moral uncertainty seriously, we should presumably also be prepared to take seriously many other forms of normative uncertainty, such as decision-theoretic uncertainty (e.g. uncertainty whether expected utility maximization is the correct norm of practical rationality), epistemological uncertainty (e.g. uncertainty whether Bayesianism is the correct account of rational credence), perhaps even uncertainty over aesthetic value.[10] While this is surely right, I will here ignore all other forms of normative uncertainty and focus exclusively on moral uncertainty.[11]

We shall shortly consider whether the standard arguments for probabilism with respect to descriptive uncertainty—representation theorem arguments, Dutch book arguments, and accuracy arguments—can also justify this doctrine with respect to moral uncertainty. Before we can do that, however, we first need to know how to represent moral claims formally. In particular, we need to make sure that it is possible to form a σ-algebra of of moral claims so that we can then define a probability function.

3.3 A FORMAL SEMANTICS FOR MORAL LANGUAGE

On the face of it, we can be uncertain about a wide range of moral claims. I may be uncertain about whether it's impermissible for me to tell a particular white lie, or about whether lying in general is impermissible. More generally still, I may be uncertain about whether some specific form of utilitarianism is the correct moral theory. Or I may be uncertain about whether modesty is a virtue, or about whether pride is a sin. Ideally, we would like to have a single framework to capture our uncertainty about all of these kinds of moral claims.

In many philosophical treatments of probabilism, the objects of credence are taken to be propositions, and these are typically analysed in accordance with the usual possible worlds semantics, which identifies each proposition with a set of possible worlds, understood as the set of worlds in which the proposition is true.[12] However, on some meta-ethical views, certain true moral claims are necessarily true, and certain false normative claims are necessarily false. For example, if utilitarianism is true, it's natural to suppose that it's necessarily true, i.e. true in every metaphysically possible world. For one thing, moral

9 See Staffel (forthcoming) for an examination of whether there are sound expressivist Dutch book and accuracy arguments for probabilism.

10 See for example MacAskill (2016a) for a discussion of decision-theoretic uncertainty.

11 For many of these domains, I believe it should be straightforward to extend the arguments for probabilism. Of course, the case of epistemological uncertainty presents a special challenge: if I'm uncertain between theories of reasoning under first-order uncertainty, it seems I may also be uncertain between theories of reasoning under second-order uncertainty and so forth, potentially leading to the same type of infinite regress we considered earlier. See Sepielli (2013) and MacAskill (2016a) for discussion of the general issue in the context of other types of normative uncertainty.

12 See Chalmers (2011) for discussion of related issues.

theories should allow us to engage in counterfactual reasoning: they should allow us to say that if the world were so and so, then this or that would be the thing to do. For another, it is often assumed that the moral supervenes on the natural, so that whenever there is a moral difference between two cases there must also be a natural difference. But this means that every moral property will be necessarily coextensive with some set of natural properties. As a result, sets of possible worlds are poorly suited for the role as objects of credence in the case of moral uncertainty, because they do not allow us to model agents who have distinct credences in different necessary truths, and distinct credences in different necessary falsehoods. All necessarily true moral claims will correspond to the same set of possible worlds (namely the set of all worlds), and all necessarily false moral claims will also correspond to the same set (namely the empty set).

The same problem arises for attempts to model failures of logical omniscience. In that context, some have proposed the use of impossible worlds to represent agents who have distinct credences in different truths of logic (e.g. Hintikka 1975). Others have suggested that such cases can be dealt with by letting sentences rather than propositions be the objects of credence (e.g. Hacking 1967 and Gaifman 2004).

Instead, however, we shall model moral claims as sets of moral theories, understood as the set of theories according to which the moral claim is true. This allows us to capture all of the different kinds of normative claims that I mentioned at the beginning of this section: the claim that it's impermissible for me to tell a particular white lie will be the set of moral theories according to which it is indeed impermissible for me to do so, the claim that lying in general is impermissible will be the set of all theories that give that verdict, and the claim that some specific form of utilitarianism is correct will be the singleton set containing that theory. Similarly, the claim that modesty is a virtue will be the set of moral theories according to which modesty is indeed a virtue. And so forth. This kind of picture is implicit in a lot of writing on moral uncertainty. For example, Lockhart (2000), Ross (2006b), Gustafsson and Torpman (2014), and MacAskill (2014) all take the objects of credence to be moral theories. Sepielli (2009:7–8) instead takes the objects of credence to be what he calls practical comparatives, i.e. propositions of the form "the balance of reasons favors doing action *A* rather than doing action *B*." However, he goes on to note that "a normative theory, on one conception at least, is just a very large practical comparative. It's a comparative of the form *Action A is better than action B, which is better than action C, which is better than action D...*"

I will argue that moral theories in fact have more structure than Sepielli allows for. In particular, I will suggest that, in addition to telling us which actions are better than which, moral theories also tell us *why* those actions are better than others. This explanatory role is a crucial aspect of moral theories. To account for it, I will borrow a framework developed by Dietrich and List (2017), which

to my knowledge is the most sophisticated formal treatment of normative theories available.[13]

3.3.1 *The Dietrich-List Framework*

We begin with a representation of the deontic content of the normative theories, i.e. their verdicts of permissibility and impermissibility. Let $\mathcal{K}$ be a set of possible choice contexts that an agent may be faced with. Each $K \in \mathcal{K}$ is a situation in which the agent has to choose among some options. For example, it might be the situation of deciding whether to order steak or vegetarian for dinner. We let $[K]$ denote the set of options in context K. For each context K, a moral theory tells us which of the available options are permissible and which are not. We represent this with a rightness function D, which maps each context K to a set $D(K) \subseteq [K]$ of permissible options. The rightness function D thus encodes the deontic content of the corresponding moral theory.

However, a moral theory is not exhausted by its deontic content. For one thing, we usually expect a moral theory to tell us *in virtue of what* the permissible actions are permissible. For another, many moral theories make more fine-grained distinctions than simply classifying actions as either permissible or impermissible. To take this into account, Dietrich and List enrich their formal framework as follows. An *option-context pair* is a pair of the form $\langle x, K \rangle$, where x is an option and K is a context. A *property* is a primitive object P that picks out a set of option-context pairs, called the *extension* of P and denoted $[P]$. Whenever a pair $\langle x, K \rangle$ is contained in $[P]$, this means that option x has property P in context K. Let $\mathcal{P}$ denote the set of those properties that may be normatively relevant. For example, we might have welfare properties, the property of being a lie, the property of being a rights-violation, and so forth.

Dietrich and List then define a *reasons structure*.

REASONS STRUCTURE A reasons structure is a pair $R = \langle N, \succeq \rangle$, consisting of (i) a *normative relevance function* N, which assigns to each context $K \in \mathcal{K}$ a set $N(K)$ of *normatively relevant* properties, and (ii) a *weighing relation* $\succeq$ over sets of properties, i.e. a binary relation over subsets of $\mathcal{P}$.

For example, in the case of a simple utilitarian theory, N will assign each context a set of utility properties, and $\succeq$ will rank sets of such properties in accordance with the criterion that more utility is better than less.

For each option x and each context K, write $\mathcal{P}(x, K)$ to denote the set of all properties of this option-context pair (among the properties in $\mathcal{P}$). The normatively relevant properties of option x in context K will be those that lie in the intersection of $\mathcal{P}(x, K)$ and $N(K)$. Let $N(x, K) = \mathcal{P}(x, K) \cap N(K)$. We can now derive a rightness function from a reasons structure by letting the permissible

13 There are some other accounts of normative theories in the literature on normative uncertainty. See for example Gustafsson and Torpman (2014: 171) and MacAskill (2014: 12).

options be the ones that are at least as choice-worthy as all other available options according to that reasons structure:

$$D(K) = \{x \in [K] : N(x, K) \succeq N(y, K) \text{ for all } y \in [K]\}.$$

With this formal framework in place, Dietrich and List go on to provide a taxonomy of various types of moral theories. I will not go through all of their distinctions, but let us consider a few just to get a better understanding of the framework. Consider a very simple moral theory such as total hedonistic utilitarianism. Intuitively, this theory says that there is only one normatively relevant property—pleasure—and that more of this property is always better. What kind of a property is pleasure? The amount of pleasure that a given action would lead to does not depend on which other actions are available. That is, pleasure is an *option property*: whether a given option-context pair has the property depends only on the option and not the context. Dietrich and List say that a moral theory is *structurally consequentialist* just in case it only ever deems option properties to be normatively relevant. So total hedonistic utilitarianism comes out as consequentialist in this sense.

Strictly speaking, total hedonistic utilitarianism recognizes not just one normatively relevant property, but many: the property of leading to amount x of pleasure, the property of leading to amount y of pleasure, etc. So we can write that for every context K, $N(K)$ is the set of all properties of the form $P_{\text{pleasure}=x}$. Furthermore, the weighing relation linearly orders singleton sets of properties of the form $\{P_{\text{pleasure}=x}\}$, so that $\{P_{\text{pleasure}=x}\} \succeq \{P_{\text{pleasure}=y}\}$ just in case $x \geq y$ But we can say that these are all properties of the same type, namely pleasure properties, and that total hedonistic utilitarianism recognizes only one type of normatively relevant property.

Total hedonistic utilitarianism is also a *monistic* (as opposed to *pluralistic* theory in that it assigns each option exactly one normatively relevant property in every context. Dietrich and List say that a theory is *teleological* if the weighing relation can be interpreted as a betterness relation. However, this betterness relation need not be consequentialist in the sense just defined. Assuming that a betterness relation is required to be transitive and reflexive, we can then say that a moral theory is teleological just in case its weighing relation is transitive and reflexive. Total hedonistic utilitarianism is clearly teleological in this sense.

3.3.2 *A General Framework for Representing Moral Claims*

With the account of moral theories in place, we can now give an account of moral claims in general. The reasons structures will play the role of possible worlds, so that each moral claim corresponds to a set of reasons structures. More specifically, let $\mathcal{R}$ be a set of reasons structures and let $\mathcal{F}$ be a σ-algebra over $\mathcal{R}$. We can now state probabilism with respect to moral uncertainty as follows:

PROBABILISM$_{\text{M}}$ A rational agent's quantitative beliefs in moral claims can be represented as a probability space $\langle \mathcal{R}, \mathcal{F}, P \rangle$.

For the most part, we shall be exclusively concerned with the agent's credences in moral claims. However, on a few occasions, particularly in section 4.2 where we discuss the place of conditionalization in a probabilistic moral epistemology we will have to consider the agent's credences in both descriptive and moral claims. In order to do so, we must appeal to a somewhat richer σ-algebra. More specifically, we will let $\Omega = \mathcal{W} \times \mathcal{R}$ be the Cartesian product of some background set $\mathcal{W}$ of possible worlds and some background set $\mathcal{R}$ of reasons structures. Then we can let $\mathcal{F}$ be a σ-algebra on Ω, and finally let P be a probability function on $\mathcal{F}$. Each proposition in the algebra will now be a set of pairs of possible worlds and reasons structures. However, we can still distinguish between descriptive claims and moral claims. Intuitively, we can understand a descriptive claim as a proposition whose moral content is the (moral) tautology, and a moral claim as a proposition whose descriptive content is the (descriptive) tautology. All other claims we can call mixed. More specifically, a proposition $D \subseteq \mathcal{W} \times \mathcal{R}$ expresses a (purely) descriptive claim just in case for every $\langle W, R\rangle \in \mathcal{W} \times \mathcal{R}$, if $\langle W, R\rangle \in D$, then for every $R' \in \mathcal{R}$, $\langle W, R'\rangle \in D$. Similarly, a proposition $M \subseteq \mathcal{W} \times \mathcal{R}$ expresses a (purely) moral claim just in case for every $\langle W, R\rangle \in \mathcal{W} \times \mathcal{R}$, if $\langle W, R\rangle \in M$, then for every $W' \in \mathcal{W}$, $\langle W', R\rangle \in M$. With this in place, the notion of the probability of a moral claim conditional on a descriptive claim that we shall draw on later is now well-defined.

This concludes our description of the formal framework. Now that we know what probabilism with respect to moral uncertainty amounts to, it's time to consider the arguments for this claim. In evaluating these arguments, I will assume the perspective of someone who is antecedently sympathetic to probabilism with respect to descriptive uncertainty. That is, I shall not have much to say about objections to probabilism with respect to moral uncertainty that apply with equal force to probabilism with respect to descriptive uncertainty. The more interesting question, I take it, is whether there are reasons to reject probabilism with respect to moral uncertainty in particular. With that in mind, let us begin with representation theorem arguments.

3.4 REPRESENTATION THEOREM ARGUMENTS

A representation theorem shows that if an agent's attitudes satisfy some set of conditions, she can be *represented* as having a probabilistic credence distribution. We can distinguish between *decision-theoretic* representation theorems, which start from a preference relation, and *epistemic* representation theorems, which start from a comparative probability relation (Briggs 2015).

In the decision-theoretic case, let $\succeq$ be a weak preference relation over some set of prospects, so that $A \succeq B$ means that A is at least as preferred as B. A decision-theoretic representation theorem shows that if $\succeq$ satisfies some particular set of constraints, then there exists a probability function P and a utility function U which together represent $\succeq$ in the sense that, for any two prospects A and B, $A \succeq B$ iff the expected value of A is at least as great as the expected value of B, where the expected values are calculated relative to P and U. Typ-

ically, the utility function will be unique up to positive linear transformation, whereas the probability function will be fully unique.

In the epistemic case, let $\trianglerighteq$ be a weak comparative probability relation over some set of prospects, so that $A \trianglerighteq B$ means that A is judged to be at least as probable as B. An epistemic representation theorem shows that if $\trianglerighteq$ satisfies some particular set of constraints, then there exists a probability function P which represents $\trianglerighteq$ in the sense that, for any two prospects A and B, $A \trianglerighteq B$ iff $P(A) \geq P(B)$.

Thus, a representation theorem argument posits that our preferences or our comparative probability judgments are rationally required to satisfy the relevant set of constraints and concludes via the representation theorem that we are rationally required to have a probabilistic credence function. Much of the philosophical work therefore consists in justifying the claim that the relevant set of constraints are indeed requirements of rationality.

Representation theorems have played a central role in decision theory and formal epistemology because the qualitative concepts embodied in the binary relations are often thought to be in some sense prior to the quantitative concepts (Bradley 2017:43). This is especially true of decision-theoretic representation theorems. Preferences manifest themselves in choice behaviour, and as a result they are more easily observable than are numerical degrees of belief and desire. They form the empirical basis for the assignment of probabilities and utilities: the qualitative attitudes provide evidence for the quantitative ones.

What, then, are preferences? According to the choice-theoretic account, claims about preferences are simply claims about choice behaviour of some kind. In the early days of revealed preference theory, preferences were identified with actual choice behaviour (Samuelson 1938). On the one hand, this makes preferences perfectly amenable to empirical investigation: we simply have to observe people's choices. On the other, it means that we can only speak of an agent's preferring ice cream over torture if she has in fact encountered a situation in which both options were available and chosen the former. More recently, some have taken preferences to be identical to hypothetical choice behaviour (Binmore 1994). This overcomes some of the difficulties of revealed preference theory, but neither version is able to account for the apparent fact that preferences can *explain* choice behaviour. My choosing pistachio ice cream over strawberry ice cream is explained by my preference for the former. But if preferences are identical to choice behaviour, then they cannot explain it.

According to the mentalist account, preferences are judgments or attitudes of a certain sort. We might initially think that they are judgments of self-interest, but further reflection reveals this to be implausible. The question of whther people are fundamentally self-interested should not be settled by stipulative fiat. As an example of a more sophisticated mentalist account, consider Hausman (2011) who proposes that preferences are total subjective comparative evaluations. They are of course comparative in that they always compare two

options (unlike, say, desire). They are subjective in that agents may have different preferences without any of them thereby being mistaken. And they are total in that they concern everything that the agent takes to have bearing on the question.

What about comparative confidence judgments? In this case, it would on the face of it seem more difficult to give a choice-theoretic account. When it comes to preference, even a mentalist agrees that there is some link or other between preferring A to B and choosing A over B, even if she denies that this link is definitional. By contrast, knowing that an agent judges A to be more likely than B does not tell us anything about how she is likely to choose, at least not in the absence of further assumptions about her preferences. But if we make the innocuous assumption that she prefers receiving £10 to not receiving £10, then we can formulate comparative confidence judgments in choice-theoretic terms by offering her to choose between a prize that pays £10 if A and nothing otherwise, and a prize that pays £10 if B and nothing otherwise.

Although I will assume that preferences and comparative confidence judgments are connected both to our choice behaviour and to various psychological states, I will not settle on a particular account. Instead, we shall now examine the prospects for formulating decision-theoretic and epistemic representation theorem arguments for the case of moral uncertainty.

3.4.1 *Decision-Theoretic Representation Theorems*

As we have seen, a decision-theoretic representation theorem takes as its starting point a binary preference relation on some set of prospects and shows that if the preference relation satisfies certain conditions, it can be represented as expected utility maximisation relative to a unique probability function and a unique (up to positive linear transformation) utility function.[14]

For our purposes, one immediate difficulty with using decision-theoretic representation theorems to establish probabilism is that preferences are a poor starting point for moral uncertainty. We cannot infer from my preference for spending £20 on dinner over donating that same £20 to charity that I believe the former to be morally better than the latter, or that I take it to have a higher expected choice-worthiness, or any other such claim. I may simply not care about morality, in which case my preferences will never reflect my moral judgments. Or perhaps I do care, but not to the point of letting morality be the sole determinant of my preferences. In neither case can my moral views be straightforwardly read off from my preferences. And this is so regardless of whether we opt for a choice-theoretic or a mentalist approach to preferences.

Hence any decision-theoretic representation theorem for moral uncertainty would have to start from some other notion than preference. And this is indeed the strategy chosen by Riedener (2015), who provides what is to my knowledge

14 See von Neumann and Morgenstern (1944/2007), Savage (1954), Jeffrey (1965), Bolker (1966), and Joyce (1999).

the only existing discussion of representation theorems in the context of moral uncertainty. More specifically, he is concerned with axiological uncertainty, i.e. uncertainty about moral value—about which options are *morally better* than which. He introduces the notion of uncertainty-relative value judgments, or u-value judgments for short. These are the betterness judgments the agent makes while taking her axiological uncertainty into account.

You might worry that u-value judgments, or whatever alternative relation we settle for, are less amenable to empirical investigation than preferences are, given that they cannot be straightforwardly read off from choice behaviour. In order to obtain an agent's u-value judgments, we have to invoke hypothetical situations: assuming you only cared about axiological considerations, or assuming that only axiological considerations were at stake, which alternative would you judge to be better? But of course, once we abandon the strictest version of the choice-theoretic account, the same is true of ordinary preferences. In order to make sense of an agent's having preferences over options she has never encountered we have to invoke her disposition to choose or the hypothetical judgments she would form, or some other such notion. Hence u-value judgments do not appear to be significantly worse off in this regard.

The standard decision-theoretic representation theorem arguments purport to establish that it is a requirement of rationality to maximise expected utility. On a natural understanding of this claim, it encompasses both theoretical and practical rationality: theoretical rationality requires us to have probabilistic degrees of belief, and practical rationality requires us to maximise expected utility relative to those probabilistic degrees of belief.

A straightforward translation of decision-theoretic representation theorem arguments from empirical uncertainty would then seem to yield the analogous conclusion that it is a requirement of rationality to maximise expected *choice-worthiness* (MEC), where the expected choice-worthiness of an option is the probability-weighted average of the choice-worthiness, or moral value, assigned to that option by each of the moral theories in which we have non-zero credence. This would give us the desired result that theoretical rationality requires degrees of belief in moral claims to be probabilistic, but it would also give us the additional result that practical rationality requires us to maximise expected choice-worthiness.

Riedener on Axiological Uncertainty

For the sake of illustration, let us briefly consider Riedener's representation theorems for axiological uncertainty. He uses the framework of *state-dependent utility theory*, which is usefully contrasted with Savage's (1954) framework. In Savage's framework, there is a set of *states of nature* and a set of *outcomes*. An *act* is a mapping from states to outcomes: it maps each state to the outcome which would result if the act were performed when that state is the actual state of nature. The agent is uncertain about which state is actual, and thereby uncertain about the outcomes of her acts. Savage's representation theorem begins

from a preference relation over acts and constructs a utility function defined over outcomes and a probability distribution over states. By contrast, in state-dependent utility theory, the utility function is defined over state-outcome pairs rather than over outcomes alone (Drèze and Rustichini, 2004). Riedener proposes state-dependent utility theory as a natural framework for axiological uncertainty. In this application, the states of nature are axiologies, and utility corresponds to u-value. Since different axiologies assign different values to empirical outcomes, the u-value of an outcome will depend on which state it occurs in. And this makes state-dependent utility theory a natural framework.

Riedener provides two representation theorems for axiological uncertainty. These theorems show under what conditions the u-value relation can be represented as maximising expected axiological value. The first theorem assumes that a probability distribution over axiologies is given exogenously. Hence it clearly cannot be used to underwrite an argument for probabilism with respect to axiological uncertainty. But it will nevertheless be useful to briefly describe it before we move on to the second theorem. This second theorem does not assume that a probability distribution over axiologies is given exogenously and is therefore more promising as an argument for probabilism.

To begin, an axiology is a binary relation over some set $\mathcal{O}$ of options. These options are probability distributions (lotteries) over (descriptive) states of the world. It is assumed that all axiologies under consideration satisfy the von Neumann-Morgenstern axioms. That is, for any $A, B, C \in \mathcal{O}$:

1. TRANSITIVITY If $A \succeq B$ and $B \succeq C$, then $A \succeq C$.
2. COMPLETENESS $A \succeq B$ or $B \succeq A$.
3. INDEPENDENCE If $A \succ B$ then $pA + (1-p)C \succ pB + (1-p)C$ for any C and any $p \in (0, 1]$.
4. CONTINUITY If $A \succ B$ and $B \succ C$, then there exist $p, q \in (0, 1)$ such that $pA + (1-p)C \succ B$ and $B \succ qA + (1-q)C$.

For his first result, which draws on Karni and Schmeidler (2016), the u-value relation is defined over a somewhat more complex set of options $\mathcal{Q}$ than are the axiologies themselves. More specifically, the options are probability distributions over *pairs* of states of the world and axiologies. This means that we are taking into account both descriptive and axiological uncertainty.

Here is the result: the u-value relation $\succeq_U$ over $\mathcal{Q}$ can be represented as maximising expected axiological value just in case it satisfies the von Neumann-Morgenstern axioms with respect to $\mathcal{Q}$ (Riedener 2015, p. 47). As I already mentioned, this result cannot be used to support an argument for probabilism. Let us therefore move on to the second result.

Although this result does not assume that we have a probability distribution over axiologies, it does assume that we have one over (descriptive) outcomes. More specifically, an option is set of pairs of, on the one hand, an axiology,

and on the other, a probability distribution over outcomes. An option contains one such pair for each axiology under consideration. We can interpret these pairs as saying: "if axiology T_i is true, then the probability distribution over outcomes is such and such." Of course, under normal circumstances the probability distribution over outcomes will be independent of which axiology is true. But the construction gives us a way of incorporating axiologies into the options without having to specify a probability distribution over them. So for the second result, the u-value relation is defined over a set $\mathcal{K}$ of options of this sort.

Again, it is assumed that the axiologies under consideration satisfy the von Neumann-Morgenstern axioms with respect to $\mathcal{O}$, and that the u-value relation satisfies them with respect to $\mathcal{K}$. However, since we are no longer assuming a probability distribution over axiologies, these conditions are not sufficient for a representation theorem. In order to get around this, Riedener supposes that the u-value relation is defined not just over $\mathcal{K}$, but also over the set $\mathcal{Q}$ that was used in the previous result. This way, the agent can form conditional u-value judgments of the form: "if the probability distribution over axiologies were P, I would judge that A is u-better than B." Roughly, the idea is now as follows: by considering the agent's conditional u-value judgments, we hold the probability distribution fixed, and can therefore determine her utility function in the same way we did with the previous theorem. Having determined her utility function, we can then use her unconditional u-value judgments ("actually, I judge that A is u-better than B") to determine her probability distribution over axiologies. Hence with this assumption in place, Riedener is able to provide a representation theorem.

Of course, by considering the agent's conditional judgments, we still had to take some notion of probability as a given. So, in what sense could this result underwrite an argument for probabilism with respect to axiologies? Riedener takes himself to be engaged in the project of giving an explication of the notion of subjective credence in axiologies. But nothing has been said so far about how the probability distributions that the agent is considering in her conditional judgments should be interpreted. If we interpret them as the agent's own subjective probabilities, you might suspect that the explication of subjective credence in axiologies becomes circular. But if we interpret them as *objective* probabilities, there is no risk of circularity. Of course, the notion of an axiology having an objective probability sounds rather outlandish. But Riedener asks us to imagine a scenario in which God used a randomising device to determine the true axiology. In order for the construction to work, we don't have to find such a scenario plausible. We only have to believe that it is conceptually coherent. If it is, then these probability distributions over axiologies can play their intended role in conditional u-value judgments, and the theorem is able to underwrite an argument for probabilism with respect to axiological uncertainty. Clearly, much more can be said here.

Can Riedener's representation theorem provide us with an argument for probabilism with respect to moral uncertainty? As it turns out, his result is limited

in a number of ways.[15] The first thing to note is that the result only pertains to axiological uncertainty (i.e. uncertainty over which states of affairs are morally better than others), and therefore says nothing about how other types of moral uncertainty, such as deontic uncertainty (i.e. uncertainty over which acts are permissible), should be represented. So even if the argument is otherwise successful, it can only establish probabilism with respect to a specific type of moral uncertainty. Secondly, the assumption that all axiologies under consideration satisfy the von Neumann-Morgenstern axioms arguably rule out some plausible axiologies, or at any rate axiologies that agents may reasonably assign non-zero credence. For example, by requiring that axiologies be complete, we rule out axiologies that posit incommensurable values. Furthermore, the continuity condition rules out axiologies that assign infinite value to some outcome and finite value to others, as well as ones according to which some kinds of values lexically dominate others. So even if we find Riedener's argument persuasive, it only establishes probabilism with respect to a particular class of axiologies.

Intertheoretic Comparisons of Value

From my perspective, one potential further downside of decision-theoretic representation theorem arguments in general is that they may prove too much. As we saw earlier, these arguments purport to establish both that theoretical rationality requires us to have probabilistic credences and that practical rationality requires us to maximise expected choice-worthiness (MEC). Given that my focus here is on whether probabilism holds for moral uncertainty, it would be preferable if we could answer that question without also having to take a stance on the question of how to make decisions, or how to evaluate options, in light of such uncertainty. Additionally, the claim that practical rationality requires us to maximise expected choice-worthiness is much more controversial than the claim that theoretical rationality requires us to have probabilistic credences. There are several accounts of decision-making under moral uncertainty which assume that agents have probabilistic credences in moral claims and yet recommend a different procedure than that of maximising expected choice-worthiness, such as acting in accordance with the moral theory in which one has highest credence (Gracely 1996, Gustafsson and Torpman 2014). Furthermore, in order for the injunction to maximise expected choice-worthiness to even be well-defined, we must assume that so-called *intertheoretic comparisons of value* are possible, because the choice-worthiness function tells us how an option's value according to one moral theory compares to its value according to another. But it's not at all clear what could ground these comparisons, and indeed, scepticism about intertheoretic value comparisons is one common motivation for those who propose an alternative account of decision-making under moral uncertainty. Intertheoretic comparisons of value are comparisons of the form:

INTERTHEORETIC COMPARISONS The difference in choice-worthiness between A and B according to moral theory T_1 is greater than the difference in choice-worthiness between C and D according to moral theory T_2.

15 I should note that Riedener himself is keenly aware of the limitations of the result.

Their name highlights the analogy with interpersonal comparisons of well-being. And just as with interpersonal comparisons of well-being, the intelligibility of intertheoretic comparisons of value has been rather controversial.

In order for MEC to be a well-defined decision rule, such comparisons must meaningful. To see why, consider the following simple case. Say that your credence is split between two theories, T_1 and T_2, and that the two available options are A and B. Let U_1 and U_2 be the choice-worthiness functions of T_1 and T_2 respectively. The expected choice-worthiness of these two options is given as:

$$EC(A) = 0.5U_1(A) + 0.5U_2(A)$$
$$EC(B) = 0.5U_1(B) + 0.5U_2(B)$$

MEC now tells us to pick whichever option has the highest expected choice-worthiness. But in order for these expressions to be well-defined, intertheoretic comparisons must be possible. We must be able to put the two theories in a common scale, so to speak, so that if $U_1(A) > U_2(A)$, this means that A is more choice-worthy according to T_1 than according to T_2. Following Gustafsson and Torpman (2014), let us distinguish between the following views:

NON-COMPARATIVISM Intertheoretic comparisons are never possible (Gustafsson and Torpman 2014).

WEAK COMPARATIVISM Intertheoretic comparisons are sometimes possible (Ross 2006, MacAskill 2014).

STRONG COMPARATIVISM Intertheoretic comparisons are always possible (Lockhart 2000, Sepielli 2010).

Why be sceptical of intertheoretic comparisons of value? I will first give two reasons to think that there are at least some classes of moral theories for which intertheoretic comparisons are not possible. If this is correct, we should reject strong comparativism. Then I will consider some reasons for thinking that they are *never* possible. If this is correct, we should reject even weak comparativism.

The first reason for thinking that intertheoretic comparisons are not always possible has to do with the structure of moral theories. Let $\mathcal{O} = \{O_1, O_2, \ldots, O_n\}$ be some set of options. The job of a moral theory is to tell us how good, or choice-worthy, these options are. One way for it to do so is to provide a ranking of those options, say $O_1 \succ O_2 \succ \cdots \succ O_n$. If this is all the information a moral theory provides us with, we will say that it is *merely ordinal*. A merely ordinal moral theory does not allow us to say anything about the magnitude of differences in choice-worthiness: it does not allow us to say, for example, that the difference in choice-worthiness between O_1 and O_2 is twice as large as that between O_3 and O_4. If a moral theory is merely ordinal, then any numerical utility function U such that, for all O_i and $O_j \in \mathcal{O}$, $U(O_i) > U(O_j)$ just in case $O_i \succ O_j$ will be just as good a representation as any other.

A moral theory is measurable on an *interval scale* if it is meaningful to say that the difference in choice-worthiness between O_1 and O_2 is twice as large as that

between O_3 and O_4. If the theory is measurable on an interval scale and U_1 is a numerical representation of it, then so is any U_2 such that for all $O \in \mathcal{O}$, $U_2(O) = kU_1(O) + m$, for $k > 0$.

In order for intertheoretic comparisons of value to be possible, all of the relevant moral theories must be measurable on an interval scale. If a moral theory is merely ordinal, it does not make sense to speak of how large the difference in choice-worthiness is between two options. But it is precisely claims of this form that we need for intertheoretic comparisons.

Some moral theories are naturally thought of as being measurable on an interval scale. For example, it is natural to think of utilitarianism as saying that (all else equal) the difference in choice-worthiness between saving two lives and saving no lives is twice as large as that between saving one life and saving no lives. If necessary, the relevant interval scale can be obtained by paying attention to how the moral theory orders prospects that involve empirical uncertainty: for example, utilitarianism will say that the option of saving one life with certainty is just as good as the option with one-half probability of saving no lives and one-half probability of saving two lives.

But not all moral theories are amenable to such a treatment. For example, many deontological theories are arguably best thought of as being merely ordinal. On Kant's view, murder is worse than lying, which in turn is worse than not helping a stranger in need. But it does not seem to make much sense to ask Kant whether the difference between murder and lying is smaller or greater than that between lying and not helping a stranger in need (MacAskill 2014:55). Relatedly, according to *absolutist* moral theories, some acts are always wrong, no matter what the consequences. Such theories are also naturally thought of as having a merely ordinal structure. Hence there seems to be many influential moral theories that lack the structure necessary for intertheoretic comparisons of value to be possible. Since a decision-theoretic representation theorem for moral uncertainty will imply that all theories under consideration are intertheoretically comparable, such a representation theorem cannot be given for an agent for whom merely ordinal theories are a live option. Therefore, it cannot be used to establish probabilism with respect to merely ordinal theories.

The second reason for thinking that intertheoretic comparisons of value are not always possible also has to do with the structure of moral theories. Some moral theories posit *incommensurable* values. For example, a theory might say that although pleasure and beauty are both valuable, they cannot be compared: we cannot say whether an outcome with some given amount of pleasure is better or worse than an outcome with some other given amount of beauty. This incommensurability can be either weak or strong. If it is weak, then there are at least some amounts for which the two are incomparable. If it is strong, then they are never comparable. Both weak and strong incommensurability spell trouble for intertheoretic comparisons of value, because such comparisons require that moral theories tell us, for any pair of options, how large the difference in choice-worthiness is between them. Therefore, a decision-theoretic

representation theorem cannot be given for an agent for whom moral theories that posit incommensurable values are a live option.

We have now seen two types of moral theories for which intertheoretic comparisons are not always possible. Moreover, it's not as though these are obscure moral theories: they include some of the most prominent ones. I therefore take it that we have good reason to reject strong comparativism.

But suppose now that an agent is only considering theories that are measurable on an interval scale. Do we still have any reason to think that intertheoretic comparisons of value are impossible?

Some argue for the claim that intertheoretic comparisons are sometimes possible by pointing to particular cases in which we would intuitively judge that such comparisons are meaningful. For example, MacAskill and Ord (forthcoming) present two such cases. The first class of cases are ones where the two theories under consideration are very similar. As an example, MacAskill and Ord consider the view that Singer (1972) presents in "Famine, Affluence, and Morality." There, he proposes a modification of common-sense ethics according to which those of us who are wealthy have stronger obligations to the global poor than we may have otherwise thought. Consider now the claim "It's more important to give to Oxfam on Singer's view than it is on the common-sense view." This claim clearly seems true. Moreover, it is a claim about intertheoretic comparisons. The second class of cases are ones where the two theories under consideration only differ with respect to the extension of bearers of value. As an example, they consider utilitarianism, which says that the value of an outcome is given by the sum of well-being across all people, and utilitarianism*, which says that the value of an outcome is given by the sum of well-being across all people except Richard Nixon. Again, some value comparisons between these two theories clearly seem to be true: for example, the claim that the two theories agree on the value of each person except Richard Nixon.

Suppose we agree that the intertheoretic comparisons in these two cases are intelligible. How far does that get us? In both cases, it seems as though one moral theory is *defined* in terms of its relation to another moral theory: Singer's view is just like the common-sense view, except that it assigns greater importance to help the global poor, and utilitarianism* is just like utilitarianism, except that it doesn't count the well-being of Richard Nixon. It is almost as if these theories are defined by stipulating that certain intertheoretic comparisons hold. This need not in itself be problematic, but when one considers the moral theories that have been influential, it seems quite rare for one of them to be defined in terms of another in this manner. So the worry is that these intuitive comparisons will only be possible in somewhat artificial cases.

Another view, advocated by Riedener, is that intertheoretic comparisons get their meaning via the relevant representation theorem. However, he thinks that probabilistic credences also get their meaning via the relevant representa-

tion theorem. But if the conditions of the representation theorem will only be satisfied by agents who are only considering moral theories of a specific type (i.e. axiologies that satisfy the von Neumann-Morgenstern axioms), then it seems we cannot meaningfully speak of agents having probabilistic credences in moral theories that are not of that type.

Upshot

We have found good reason to reject strong comparativism, while leaving some room for weak comparativism. In light of this, what are the prospects for a decision-theoretic representation theorem argument for probabilism with respect to moral uncertainty? If non-comparativism is correct, then clearly no such argument can get off the ground. If a decision-theoretic representation theorem is to underpin a fully general argument for probabilism with respect to moral uncertainty, it is not sufficient to assume weak comparativism. Strong comparativism is required. According to weak comparativism, intertheoretic comparisons are possible for some moral theories but impossible for others. Suppose now that an agent has preferences, or u-value judgments, over moral theories of both kinds. What are the rationality constraints on these u-value judgments? If we go for the ones that that figure in a decision-theoretic representation theorem, we will of course end up with a representation of the agent as an expected utility maximiser. But such a representation implies that all theories are intertheoretically comparable. Hence it cannot be that our u-value judgments must satisfy the axioms of the representation theorem for every moral theory. In particular, they must violate at least some of them whenever the judgment concerns two incomparable theories.

In response, we might settle for a more modest conclusion: if an agent is only considering intertheoretically comparable theories, then her degrees of belief in those theories are rationally required to be probabilistic. For example, suppose that we are concerned with axiological uncertainty, and that the agent in question is only considering axiologies that satisfy the von Neumann-Morgenstern axioms. We can then argue that the u-value relation should also satisfy these axioms and then derive probabilism via Riedener's second result. If on the other the agent is also considering some incomparable theories, then we don't make any claims about what rationality requires of her degrees of belief.

However, although the argument for this more modest conclusion may be sound, it is not an especially satisfying conclusion. As we have seen, the class of incomparable theories includes several prominent moral theories. By excluding these from consideration, we seem to be giving up the game. We might hope that some other argument can underwrite probabilism with respect to credences in incomparable theories as well. But any argument that does so is likely to lend just as much support to probabilism with respect to credences in comparable theories (as we shall see when we examine Dutch book arguments and accuracy arguments later on). Hence by appealing to a separate argument to address the special case of incomparable theories, we undermine the decision-theoretic representation theorem argument's standing as an inde-

pendent argument for probabilism.

So, the upshot is that decision-theoretic representation theorems cannot underwrite an argument for probabilism with respect to moral uncertainty in general. Therefore, insofar as you accept probabilism with respect to descriptive uncertainty on the basis of a decision-theoretic representation theorem argument, you have less reason to accept it with respect to moral uncertainty.

3.4.2 *Epistemic Representation Theorems*

As you will recall, epistemic RTAs take as their starting point a notion of comparative confidence, rather than preferences (like the standard decision-theoretic RTAs) or u-value judgments (like Riedener's repurposing of standard decision-theoretic RTAs). Again, we begin with some set Ω and a σ-algebra $\mathcal{F}$ on Ω. The comparative confidence relation $\trianglerighteq$ is a binary relation on $\mathcal{F}$. We will gloss $A \trianglerighteq B$ as "A is as least as likely as B." The general strategy of the argument is the same as for decision-theoretic RTAs. We impose some purported rationality constraints on the relevant binary relation, and then show that it can be given a certain type of numerical representation just in case it satisfies those rationality constraints. In the case of decision-theoretic RTAs, the numerical representation is an expected utility representation with respect to relevantly unique probability and utility functions. But in the case of epistemic RTAs, the numerical representation is just the probability function alone. In dispensing with the utility function, epistemic RTAs avoid what I took to be perhaps the main pitfall for decision-theoretic RTAs, namely that a fully general decision-theoretic RTA for probabilism with respect to moral uncertainty would necessarily commit us to strong comparativism, i.e. it would force us to say that intertheoretic comparisons of differences in choice-worthiness are always possible. Unlike the preference relation, the comparative confidence relation is fully epistemic, and therefore has no need for value comparisons.

The Comparative Confidence Relation

What does it mean to judge that A is at least as likely as B? And how can we find out whether an agent makes this judgment or not? We know how to determine whether she *prefers* A to B: that she would (in suitable circumstances) choose one over the other. Although their link to choice is not definitional, preferences do nevertheless manifest themselves in the corresponding choice behaviour with sufficient regularity. This makes preferences amenable to empirical investigation. Because the link between comparative confidence judgments and choice behaviour must necessarily be weaker, they may seem to be proportionally less amenable to empirical investigation.

Consider the most natural idea for a test. In order to determine whether or not an agent judges that A is more likely than B, we offer them a desirable prize and ask them to choose between either receiving this prize if A is true or receiving that same prize is B is true. If they choose the former, we can conclude that they judge A to be more likely than B. I think this is a fine test.

But notice that it does require us to make some assumptions about the agent's preferences. In particular, we have to assume that the agent prefers receiving the prize to not receiving it, and that this preference remains the same no matter which one (if any) of A and B turns out to be true. So a natural version of the simple choice-theoretic account would identify an agent's comparative confidence judgments with her choice behaviour in situations of this kind. And more nuanced choice-theoretic accounts would appeal to her dispositions to choose in such situations, or her hypothetical choice behaviour.

So on the face of it, epistemic representation theorems for probabilism with respect to moral uncertainty would seem to have more going for them than the decision-theoretic ones. They avoid having to make intertheoretic comparisons. And comparative confidence judgments appear to be just as amenable to empirical investigation, at least provided some innocent assumptions about how they can be revealed in choice behaviour. However, given that comparative confidence is a cognitive attitude (unlike preference which is an evaluative attitude), you might wonder whether we have to assume cognitivism in order to make sense of an agent's making comparative confidence judgments about moral claims. Let us see how the expressivist framework we have looked at deals with this. As you will recall, Schroeder introduces the notion of being for. To capture moral uncertainty, we introduce the notion of degrees of being for. So we would then read $A \trianglerighteq B$ as saying that the agent is more for a than she is for b, where a and b are suitable analyses of what A and B respectively say that the speaker is for. Therefore, if you are happy to allow that expressivists can account for moral uncertainty more broadly, the comparative confidence relation should not pose any special problems.

The Axioms of Qualitative Probability

In the first half of the 20th century, several probabilists were studying the comparative confidence relation. Many took it to be in some sense more fundamental than quantitative probability (Koopman 1940a, 1940b; Good 1950; de Finetti 1951, Savage 1954). In particular, they thought that comparative confidence judgments could play an important part in explicating the concept of degrees of belief.[16] What conditions should a rational comparative confidence relation satisfy? De Finetti (1931) proposed the following axioms on $\trianglerighteq$.

QUALITATIVE PROBABILITY The comparative confidence relation $\trianglerighteq$ on $\mathcal{F}$ is said to be a qualitative probability relation if for any $A, B, C \in \mathcal{F}$:

- A1. TRANSITIVITY If $A \trianglerighteq B$ and $B \trianglerighteq C$, then $A \trianglerighteq C$.
- A2. COMPLETENESS Either $A \trianglerighteq B$ or $B \trianglerighteq A$.
- A3. QUALITATIVE ADDITIVITY If $A \cap C = B \cap C = \emptyset$, then $A \trianglerighteq B$ iff $A \cup C \trianglerighteq B \cup C$.
- A4. NORMALITY $\Omega \trianglerighteq A \trianglerighteq \emptyset$.

16 For an overview of the mathematical work on comparative confidence, see Fine (1973) and Fishburn (1986).

De Finetti (1951) conjectured that these conditions were necessary and sufficient for probabilistic representability. However, Kraft et al (1959) showed this to be false: it is possible to construct a comparative confidence relation which satisfies A1-A4 and yet does not admit of a probabilistic representation. But although the conditions are not sufficient, they are necessary. We should therefore examine how plausible they are as rationality requirements on comparative confidence judgments. And in particular, we should examine whether they become any more or any less plausible when the domain of the comparative confidence judgments changes from descriptive states of the world to moral claims and moral theories. Most of these are quite straightforward. I shall have most to say about completeness, and will therefore save it for last. After having evaluated these axioms, we will look at two epistemic representation theorems, due to Scott (1964) and Villegas (1964) respectively. We shall also consider a related result for imprecise probabilism (Bradley 2017).

Transitivity
It seems clear that any rational comparative probability relation should be transitive. Moreover, the shift from descriptive uncertainty to empirical uncertainty does not seem to have any bearing on its status as a requirement of rationality.[17]

Qualitative Additivity
This axiom says that if two propositions A and B are both logically inconsistent with C, then A is more credible than B just in case the disjunction of A and C is more credible than the disjunction of B and C. This should not be controversial.

Normality
Normality simply says that no proposition is strictly more probable than the tautology, and that every proposition is at least as probable as the contradiction. Again, this should not be controversial as a requirement of rationality.

Completeness
Completeness demands that agents be fully opinionated: for any pair of propositions, they must either judge one to be more probable than the other, or judge them both to be equally probable. Many see completeness as mainly a structural assumption rather than a rationality condition in its own right. Roughly speaking, it is needed in order to ensure that her comparative confidence judgments are sufficiently rich for probabilistic representability.

Of the four axioms, A2 is the only one that makes an existential requirement: it demands that the agent form an opinion on everything. By contrast, A1 and A2 are both conditional requirements: they say that if the comparative confidence relation satisfies certain conditions, it must also satisfy certain other conditions. Similarly, A4 says that if the agent forms a judgment about A, then she must judge it to be more probable than the contradiction and less probable

17 Some of the reasons people give for abandoning completeness are also reasons to weaken transitivity somewhat (Bradley 2017:234). I will not go into detail, but the relevant weakening is known as *Suzumura consistency* (Bossert and Suzumura 2010).

than the tautology.

An agent who does not feel compelled to make up her mind does not seem to be guilty of a failure of rationality. On these grounds, many have argued that incomplete comparative confidence judgments are rationally permissible. But others have made the stronger claim that incomplete comparative confidence judgments are sometimes rationally *required*. Typically, this argument appeals to the same evidentialist considerations that were used to motivate imprecise Bayesianism in the previous chapter, i.e. the idea that in cases where we only have limited, partial, or ambiguous evidence, completeness would require us to go beyond the information available to us in an epistemically irresponsible way (Joyce 2010). If one finds these considerations convincing in the case of descriptive uncertainty, one should presumably say the same in the case of moral uncertainty. Moreover, one may think that if anything, the evidentialist case against completeness is stronger in the case of moral uncertainty. I take it that, regardless of one's views in moral epistemology, it is at least *prima facie* plausible to think that the evidence we have for or against various moral claims is in some sense sparser than the evidence we have for or against various moral claims. This notion of sparseness can be spelt out in different ways. I will just consider one example. In physics, the empirical evidence available to us lets us determine the value of Planck's constant to at least ten significant figures. But if say a pluralist moral theory is true, according to which both utility and equality are of value, it is unlikely that we will ever be able to establish the correct conversion rate between these two goods with comparable precision. Hence if one thinks that unnecessary precision is rationally impermissible, one will find rational violations of completeness to be more frequent in the moral case.

This shows why it matters whether completeness is merely a structural assumption or not. If it is merely a structural assumption, then we can say that although completeness is not rationally required, it is rationally permissible, and an agent should therefore at least in principle be able to become fully opinionated. But if she does become fully opinionated, then the rationality assumptions entail that her comparative confidence judgments can be represented by a probability function.

The consistency concepts that we introduced in section 1.2.1 will be useful here. As a reminder, recall that a belief set $\mathcal{B}$ is

STRICTLY INCONSISTENT iff there is some $A \in \mathcal{F}$ such that $A, \neg A \in \mathcal{B}$.

LOGICALLY NON-OMNISCIENT iff there is some $A \in Cn(\mathcal{B})$ such that $A \notin \mathcal{B}$.

IMPLICITLY INCONSISTENT iff there is some A such that $A, \neg A \in Cn(\mathcal{B})$.

COHERENTLY EXTENDABLE iff $\mathcal{B}$ has some strictly consistent completion $\mathcal{C}$.[18]

18 Recall that a *completion* $\mathcal{C}$ of $\mathcal{B}$ is a maximal subset of $\mathcal{F}$ such that if $A \in \mathcal{B}$ then $A \in \mathcal{C}$, and that a subset $\mathcal{S}$ of $\mathcal{F}$ is *maximal* just in case for every $A \in \mathcal{F}$, either $A \in \mathcal{S}$ or $\neg A \in \mathcal{S}$ (or both).

Let us now apply this to the comparative confidence relation. Her state of belief will now be represented as the set $\mathcal{B}_{\trianglerighteq}$ of all comparative confidence judgments she endorses, i.e. all claims of the form $A \trianglerighteq B$. The relevant consequence operator will now be qualitative probability consequence. That is, consequence with respect to A1-A4. So $Cn(\mathcal{B}_{\trianglerighteq})$ will be the closure of her belief state under qualitative probability consequence. This lets us define all the same consistency concepts as before. The rationality requirement on comparative confidence judgments is now this: that they be coherently extendable. That is, for any $A, B \in \mathcal{F}$, she should be able to add either $A \trianglerighteq B$ or $B \trianglerighteq A$ to her set of comparative confidence judgments without becoming strictly inconsistent. Given that, standardly, a belief state $\mathcal{B}_{\trianglerighteq}$ is coherently extendable just in case it is implicitly consistent, it follows that, whether or not it is complete, any rational set of comparative confidence judgments $\mathcal{B}_{\trianglerighteq}$ must satisfy the other qualitative probability axioms. So on this view, coherent extendability is the normative core of probabilism.[19]

If we instead think that incomplete comparative confidence judgments are sometimes rationally required, then matters are not quite so straightforward. Why care about whether an agent's comparative confidence judgments are coherently extendable if the coherent extension in the form of a precise probability function is not even a rationally permissible doxastic state? However, most of those who think that incomplete comparative confidence judgments are sometimes rationally required also think that *complete* comparative confidence judgments are sometimes rationally required. That is, they think that there is some body of evidence $\mathcal{E}$ such that some precise credence function P is rationally required with respect to that evidence. For example, we can imagine that the agent's body of evidence consists only of chance propositions and that there are enough chance propositions to determine, via the correct chance-credence principle, a unique probability function. Plausibly, she should then apply this chance-credence principle. Assuming all of the chances are precise probabilities, her credence function will be a precise probability function.

With this in mind, we can appeal to a principle that is similar to coherent extendability. First, we assume that any rational agent should in principle be able to acquire some body of evidence $\mathcal{E}$ such that it is rationally permissible to be fully opinionated in light of that body of evidence. We say that a set of comparative confidence judgments $\mathcal{B}_{\trianglerighteq}$ is *evidentially extendable* just in case it has some strictly consistent completion $\mathcal{C}$ such that $\mathcal{C}$ would be rationally permissible relative to some body of evidence $\mathcal{E}$ that an agent with comparative confidence judgments $\mathcal{B}_{\trianglerighteq}$ could in principle learn. We can now run through a similar argument:

19 We can also give coherent extendability an expressivist reading. For Schroeder, it says that an agent should in principle be able to come to be more for a than she is for b, for any $A, B \in \mathcal{F}$ (where a and b, recall, are suitable analyses of whatever it is that A and B say that the speaker is for) without thereby becoming strictly inconsistent, where strict inconsistency is now understood as a property of the set of contents of the agent's *For* attitudes. This seems like a plausible requirement of rationality.

1. The set of comparative confidence judgments $\mathcal{B}_{\trianglerighteq}$ is rationally required to be evidentially extendable.

2. Typically, $\mathcal{B}_{\trianglerighteq}$ is evidentially extendable just in case it is implicitly consistent.

3. Therefore, whether $\mathcal{B}_{\trianglerighteq}$ is complete or not, it should satisfy the other axioms on qualitative probability.

In the previous chapter I argued that on at least one natural way of spelling out the evidentialist argument against completeness, it has the unacceptable implication that the agent's imprecise credences with respect to a very large class of propositions will barely change at all no matter how much (finite) evidence she receives. But there are other ways of capturing at least part of this evidentialist line of thought which do not lead to global belief inertia. In particular, we could say that, in cases where multiple probability assignments to A are all compatible with the evidence, then imprecision is rationally permissible but not rationally required. Therefore, we can reject completeness without being forced to say that global belief inertia is rationally required.

Let us now turn to the two representation theorems.

Scott's Theorem

One well-known epistemic representation theorem is due to Scott (1964). To prove this theorem, we need one further condition beyond the qualitative probability axioms. Let $\mathbf{A}^i$ be the *indicator function* of any set A, i.e. $\mathbf{A}^i(\omega) = 1$ if $\omega \in A$ and $\mathbf{A}^i(\omega) = 0$ otherwise. Scott's condition may now be stated as follows.

SCOTT'S CONDITION For all $A_0, \ldots, A_n, B_0, \ldots, B_n \in \mathcal{F}$, if $A_i \trianglerighteq B_i$ for $0 \leq i < n$, and for all $\omega \in \Omega$,

$$\mathbf{A}_0^i(\omega) + \cdots = \mathbf{A}_n^i(\omega) = \mathbf{B}_0^i(\omega) + \cdots + \mathbf{B}_n^i(\omega),$$

then $B_n \trianglerighteq A_n$.

This condition is not especially intuitive, so let's try to unpack it. First, we have two sequences $A_0, \ldots, A_n$ and $B_0, \ldots, B_n$ of the same number of propositions. The first part of the antecedent requires that the ith element of the first sequence is more probable than the ith element for every $i \neq n$. If $\mathbf{A}^i(\omega) = 1$ then the proposition A is true at ω, and if $\mathbf{A}^i(\omega) = 0$, then A is false at ω. For a given ω, the central equality states that the same number of A propositions and B propositions are true at ω. The antecedent requires this to hold for all $\omega \in \Omega$. This means that, no matter which world is actual, it is guaranteed that the same number of A propositions and B propositions will be true. Finally, Scott's condition says that for any two sequences of propositions in the algebra that satisfy these conditions, it must be the case that the final element of the second sequence is more probable than the final element of the first sequence.

Scott's condition implies both A1 and A3. Let us see how it implies transitivity. For any $A, B, C \in \mathcal{F}$, and for every $\omega \in \Omega$, we have that

$$\mathbf{A}^i(\omega) + \mathbf{B}^i(\omega) + \mathbf{C}^i(\omega) = \mathbf{B}^i(\omega) + \mathbf{C}^i(\omega) + \mathbf{A}^i(\omega).$$

Suppose now that $A \trianglerighteq B$ and $B \trianglerighteq C$. This means that the antecedent of Scott's Condition is satisfied, and therefore it follows that $A \trianglerighteq C$.

Theorem 2. (Scott 1964) Let Ω be a finite set and let $\mathcal{F}$ be a σ-algebra on Ω. If the comparative confidence relation $\trianglerighteq$ on $\mathcal{F}$ satisfies A2, A4, and SC, then there is a probability function P such that for all $A, B \in \mathcal{F}$, $P(A) \geq P(B)$ iff $A \trianglerighteq B$.

Scott's theorem does not guarantee that the probability function in question is unique. This should not be surprising. Given that Ω was assumed to be finite, there will only be a finite number of possible qualitative probability relations $\trianglerighteq$ on $\mathcal{F}$. Therefore, each qualitative probability relation will be represented by some set of probability functions rather than a unique one.

Scott's theorem gives us necessary and sufficient conditions for when a comparative confidence relation can be given a probabilistic representation in the finite case. However, the axiom he appeals to is not especially intuitive, and few have therefore undertaken to defend it as a rationality constraints. There are other results using more intelligible axioms that give us only sufficient conditions (e.g. Suppes 1969). But we shall now consider the infinite case instead.

Villegas's Theorem

For our next result, we will require both the event space Ω and the σ-algebra $\mathcal{F}$ to be uncountably infinite. As before, we let $\trianglerighteq$ be a qualitative probability relation in de Finetti's sense (i.e. a binary relation satisfying A1-A4). An *atom* of a σ-algebra $\mathcal{F}$ is a proposition $A \in \mathcal{F}$ such that $A \neq \emptyset$ and for every $B \in \mathcal{F}$, if $B \subseteq A$, then $B = A$. All finite σ-algebras have atoms. An algebra is *atomless* iff it has no atoms. We say that σ-algebra $\mathcal{F}$ is *complete* iff every $\mathcal{S} \subseteq \mathcal{F}$ has a lower and upper bound with respect to set inclusion, i.e. elements $B_*, B^* \in \mathcal{S}$ such that for all $A \in \mathcal{S}$, $B_* \subseteq A$ and $A \subseteq B^*$.

Villegas's theorem requires the algebra to be both complete and atomless. It also requires one further condition on the comparative confidence relation $\trianglerighteq$, namely that it be continuous, in the following sense. Let $\mathcal{A} = \{A_1, A_2, \dots\}$ be a countable set such that $A_1 \supseteq A_2 \supseteq \cdots$. Suppose that $B \trianglerighteq A_i$ and $A_i \trianglerighteq C$ for all i. The comparative confidence relation $\trianglerighteq$ is *continuous* if for every such $\mathcal{A} \in \mathcal{F}$, $B \trianglerighteq \bigcup \mathcal{A}$ and $\bigcup \mathcal{A} \trianglerighteq C$.

Theorem 3. (Villegas 1964) Let $\mathcal{F}$ be a complete, atomless σ-algebra on Ω. Let $\trianglerighteq$ be a continuous qualitative probability relation on $\mathcal{F}$. Then there exists a unique (countably additive) probability function P on $\mathcal{F}$ such that $A \trianglerighteq B$ just in case $P(A) \geq P(B)$.

The additional conditions for Villegas are essentially richness conditions that are necessary to ensure that the probabilistic representation is unique. Therefore, armed with the notion of coherent extendability, we can take Villegas's

theorem to underwrite an argument for imprecise probabilism. More specifically, consider the following condition on $\trianglerighteq$:

WEAK AXIOM OF CONSISTENCY The relation $\trianglerighteq$ has a minimal coherent extension on $\mathcal{F}$ that is a continuous qualitative probability relation.

We can now obtain the following:

Theorem 4. (Bradley 2017:236) Let $\trianglerighteq$ be a non-trivial comparative confidence relation on $\mathcal{F}$ that satisfies the weak axiom of consistency; then there exists a maximal set of probability functions $\mathcal{P} = \{P_1, P_2, \dots\}$ that rationalises $\trianglerighteq$ in the sense that for all $A, B \in \mathcal{F}$,

$$A \trianglerighteq B \Leftrightarrow \forall P \in \mathcal{P}, P(A) \geq P(B).$$

However, in order for these arguments to underwrite an argument for imprecise probabilism with respect to moral uncertainty in particular, we must first verify that the richness conditions make sense in this domain. That is, can we ensure that a σ-algebra $\mathcal{F}$ of sets of reasons structures is both atomless and complete? Begin with the former. First we need to ensure that the background set Ω of reasons structure over which $\mathcal{F}$ is defined is countably infinite. The most natural way to do this, I take it, is to let the number of choice contexts over which the reasons structures are defined be countably infinite. This guarantees that the background set Ω of reasons structures is countably infinite, and therefore that we can form an uncountably infinite σ-algebra $\mathcal{F}$ over Ω.

We now wish to impose the constraint that $\mathcal{F}$ contain no atoms. That is, we wish to impose the requirement that for every $A \in \mathcal{F}$ there is a $B \in \mathcal{F}$ such that $B \subset A$. This means that no singleton set of an individual reasons structure $\mathcal{R}$ can be an element of $\mathcal{F}$; that is, the agent is never allowed to believe a fully specified moral theory. Is this a problem? Arguably not, given the countably infinite number of choice contexts. Moreover, for a given set of choice contexts of interest, we can always make a moral claim as precise as we want with respect to those ones. So the assumption that the algebra be atomless seems okay. At the very least, it seems that it should be rationally permissible for an agent to have their probability function be defined over an atomless algebra.

In conclusion, epistemic representation theorem arguments provide a strong case for thinking that precise probabilistic credences in moral claims are rationally permissible, and that imprecise probabilistic credences in moral claims are rationally required.

3.5 DUTCH BOOK ARGUMENTS

Suppose that you are in the business of placing bets, and suppose further that your credences match your betting prices: your credence in A is x just in case you take £x to be the value of a bet that pays £1 if A and nothing otherwise. Your goal in placing bets is, naturally, to make money. As such, one thing you certainly wish to avoid is accepting a set of bets that jointly entail a sure loss.

Such a combination of bets is called a *Dutch book*, and anyone who accepts a Dutch book is guaranteed to lose money no matter what the actual state of the world turns out to be. Therefore, the cunning bettor who offers a Dutch book does not have to rely on any superior knowledge or information to turn a profit: she is simply exploiting the internal structure of your betting prices.

What can you do to avoid Dutch books? According to the *Dutch book theorem*, if your betting prices violate the probability calculus, you are vulnerable to a Dutch book. And according to the *converse Dutch book theorem*, if your betting prices satisfy the probability calculus, you are not vulnerable to a Dutch book. These two theorems form the mathematical centrepiece of the *Dutch book argument* that seeks to establish probabilism on the basis of the irrationality of sure losses. Here is a sketch of the argument.[20]

1. Your credences match your betting prices.
2. A Dutch book can be made against you just in case your betting prices violate the probability calculus.
3. If a Dutch book can be made against you, then you are susceptible to sure losses.
4. If you are susceptible to sure losses, then you are irrational.
5. Therefore, if your credences violate the probability calculus, then you are irrational.

Of course, this argument may be resisted in a number of ways. The second premise is simply a piece of sound mathematics, but objections have been raised for all other premises. I will not attempt an exhaustive survey, but simply sketch the main ones so that we can examine whether they apply with equal force when the Dutch book argument is intended to establish probabilism for moral rather than descriptive uncertainty.

The first premise assumes that you value money linearly. This is clearly not realistic. If you are £2 short of a tube ticket, the value you assign to £1 will be much smaller than half the value you assign to £2. We might attempt to fix this by replacing money with utility, as it is much more plausible to think that we value utility linearly. However, utility is usually defined via representation theorems, and as we saw in the previous section, those representation theorems already provide us with probabilism on their own. Hence in replacing money with utility, the Dutch book argument may lose its force as an independent argument for probabilism.

The third premise assumes the package principle that the value you assign to a combination of bets is the sum of the values you assign to the individual bets. One way for this principle to fail is if you don't value money linearly: if you are £2 short of a tube ticket, then the value you assign to two bets that

20 See Vineberg (2016) and Hájek (2008) for introductions to Dutch book arguments.

each pay out £1 will be much greater than the sum of the values you assign to them considered separately. Another way for it to fail is if you believe that your placing one bet gives you evidence about the outcome of another bet in the package. For then the act of placing the first bet will change your credence in, and thereby your fair price for, the second bet.

The fourth premise stands in need of clarification: what does it mean for a loss to be sure? Suppose that in the early 19th century you were offered to bet on the proposition that water is not H_2O. If you accept, you are guaranteed to lose as a matter of metaphysical necessity. You may be ignorant of chemistry, but you are certainly not irrational in the sense intended by the Dutch book argument. The Dutch book argument was meant to show that you can be exploited monetarily if your credences are internally inconsistent in a certain way. But you can have perfectly coherent probabilistic credences and yet be less than certain that water is H_2O. Hence it is not the case that just any susceptibility to sure losses makes you irrational.

3.5.1 *Dutch Book Arguments and Moral Uncertainty*

How does the Dutch book argument fare when we turn to moral uncertainty? On the face of it, the prospects may not seem so good. First, it's strange to imagine a situation in which an agent is offered to bet on the truth of a moral proposition. How could such a bet be settled? The situation is strange in another respect too: for on some meta-ethical views, certain general moral truths are necessarily true. But if so, there's a risk that the Dutch book argument for probabilism with respect to moral uncertainty proves too much by establishing that our credence in any necessary moral truth should be 1, just like the ordinary Dutch book argument establishes that our credence in any tautology should be 1. Second, the linearity assumption looks to be much more problematic in the case of moral uncertainty, because the value of a given amount of money may depend on which moral theory is correct.

However, I will argue that all of these challenges can be met. More specifically, I will argue that if you believe that a particular depragmatized account of Dutch book arguments gives us a sound case for probabilism with respect to descriptive uncertainty, then that same account also gives us probabilism for moral uncertainty.

Begin with the betting scenario. We might reject the setup on the grounds that it's hard to see how a bet on the truth of a moral proposition could be decisively settled. But of course, this is true of many bets on descriptive propositions as well. Consider for instance the proposition that the rate at which the universe expands at some particular point in the future when all intelligent life has perished will be thus and so. Naturally, a bet on such a proposition could never be settled. So if the Dutch book argument requires us to be able to decisively settle any bet, this is just as much of a problem for the descriptive case as it is for the moral case.

It might be objected that there is a difference between bets that could in principle be settled and ones that could not, with the proposition about the universe's rate of expansion falling into the former category and moral propositions falling into the latter. For the former category, at least we know which circumstances would verify the claim and which would falsify it, whereas we seem to lack even this in the latter case. However, we can imagine a situation in which you are offered to bet on some moral claim before God informs you of its truth value. Then even this bet can be decisively settled. If you find the example too outlandish, it suffices to imagine a situation in which someone merely believes that God or some similarly reliable being will inform them of the truth value of the relevant moral claim.

Linearity

The linearity assumption in the first premise of the Dutch book argument looks especially problematic in the case of moral uncertainty, because the value of a given amount of money may depend on which moral theory is correct. For example, perhaps a given amount of money is more valuable on a purely utilitarian theory than it is on a purely egalitarian theory. Of course, in order for this to be intelligible, we must assume that intertheoretic comparisons of value are at least sometimes possible. But ideally, probabilism with respect to moral uncertainty should we compatible with any position on intertheoretic comparisons. In discussing decision-theoretic representation theorem arguments, we wanted to avoid making the case for probabilism rest on the assumption of strong comparativism. Similarly, in discussing Dutch book arguments, we want to avoid making the case rest on non-comparativism. And if we don't assume non-comparativism, then intertheoretic comparisons may sometimes be possible, and it therefore follows that the value of a given amount of money may depend on which moral theory is correct.

Moreover, the response of formulating the argument in terms of utility rather than in terms of money is if anything even less compelling in the case of moral uncertainty. We saw that the standard objection to this response is that utility is typically defined via a decision-theoretic representation theorem. But a decision-theoretic representation theorem also gives us a probability function, and in replacing money with utility, the Dutch Book argument therefore loses its status as an independent argument for probabilism. The same point applies to moral uncertainty, but here the situation is even worse because as we have seen, the decision-theoretic representation theorem argument for probabilism with respect to moral uncertainty is unsound as it requires us to assume that non-comparable theories are in fact comparable.

However, in the case of Dutch book arguments for probabilism with respect to descriptive uncertainty, some have argued that a so-called *depragmatized* understanding of the argument can help us avoid many of the objections. I will argue that the same can be done for moral uncertainty. Christensen (2004) begins by considering what he calls *simple agents* Simple agents value money linearly, and do not value anything else. Clearly, such agents are possible, and

their values rationally permissible. Christensen suggests the following principles for simple agents:

SANCTIONING A simple agent's degrees of belief sanction as fair monetary bets at odds matching his degrees of belief.

BET DEFECTIVENESS For a simple agent, a set of bets that is logically guaranteed to leave him monetarily worse off is rationally defective.

BELIEF DEFECTIVENESS If a simple agent's beliefs sanction as fair each of a set of bets, and that set of bets is rationally defective, then the agent's beliefs are rationally defective.

According to the first principle, an agent's degrees of beliefs may sanction bets as fair without her thereby being disposed to bet or evaluate bets in the sanctioned manner. Christensen intends for the connection to be purely normative. This means that if the agent's if the agent's degrees of belief are non-probabilistic, she is not necessarily at risk of suffering financial consequences. The second principle essentially says that it would be rationally defective of a simple agent to accept a set of bets that is logically guaranteed to give her less of what she wants. Finally, the third principle forges a link between rationally defective bets and rationally defective beliefs. Together with the Dutch Book Theorem, these principles give us the following conclusion:

SIMPLE AGENT PROBABILISM If a simple agent's degrees of belief violate the probability axioms, they are rationally defective.

But what bearing does this have on agents who are not simple? The point of the Dutch book argument is not dependent on the particular preferences of a simple agent. As Christensen puts it, "there is no way the world could turn out that would make the set of bets work out well—or even neutrally—for the agent." And the assumption that the agent values money linearly plays no essential role in establishing this conclusion. This also allows Christensen to address the charge that Dutch book arguments fail to establish probabilism as a distinctively epistemic claim: the fact that an agent who bets in accordance with her beliefs is logically guaranteed to lose money reveals that those beliefs are themselves rationally defective.

Much more can be said about Christensen's argument, but I will not undertake a detailed evaluation.[21] Instead, let us see what all of this implies for moral uncertainty. Christensen's framework can help us address the worry that the value of a given amount of money may depend on which moral theory is correct. Again, we restrict our attention to simple agents, i.e. agents who value money linearly and who do not value anything else. Again, we are concerned with their betting behaviour. But of course now the bets concern moral claims rather than descriptive claims. Now it could well be that the (moral) value of a given amount of money depends on which moral theory is correct. But by

21 See Maher (1997) for criticism.

hypothesis, a simple agent is indifferent to such concerns, and they will therefore have no effect on her betting behaviour.

We are therefore, in a sense, making the opposite assumption of one we made in the context of decision-theoretic representation theorem arguments. There we assumed that the relevant attitudes—comparative u-value judgments—*only* took moral considerations into account. Here we are instead assuming that the betting behaviour is entirely insensitive to moral considerations. Therefore, insofar as we find Christensen's principles for simple agents compelling in the descriptive case, we should also find them compelling in the moral case. Consequently, we get the conclusion of simple agent probabilism with respect to moral uncertainty. And again, the Dutch book argument is not dependent on the particular preferences of a simple agent.

It may be objected that it is irrational for an agent to bet in a way that is contrary to her moral convictions.[22] Suppose that this is right. How then should an agent bet in light of her moral uncertainty? Suppose first that the agent has full credence in some moral theory T which values money non-linearly. Given that she has full credence in T, she will presumably accept any bet on the truth of T (assuming at least that T values money positively and in a monotonically increasing way, that there is no moral prohibition on gambling, etc.). Therefore, in such a case we can read off her credences from her betting behaviour even though she has positive credence in a moral theory that values money non-linearly.

Suppose next that the agent has 0.5 credence in a linear theory T_1 and 0.5 credence in a non-linear theory T_2. The agent now wishes to let these credences affect her betting behaviour. How should she do so? She is no longer in a position to be guided by a moral theory (unless she is first willing to make what would by her own lights be an entirely arbitrary choice between T_1 and T_2). Plausibly, then, her betting behaviour should somehow be sensitive to both of these theories. But how exactly?

Assume for now that she regards her own betting behaviour as evidentially irrelevant with respect to T_1 and T_2. If both T_1 and T_2 evaluate a bet as favourable, then clearly so should the agent herself. And similarly if both T_1 and T_2 evaluate it as unfavourable. But what if one evaluates it as favourable and the other as unfavourable? We could say that it is permissible (but not obligatory) to bet in such situations. If we don't think that intertheoretic comparisons are possible, we would presumably have to say something like this.

On the other hand, if we do think that intertheoretic comparisons are possible, we could attempt a more nuanced Maximise Expected Choice-Worthiness solution. In the former case, we would not be able to pin down the agent's credences by observing her betting behaviour. In the latter case, the utility function in question would presumably come from a decision-theoretic represen-

22 For example, see Nissan-Rozen (2015) and Bradley and Stefánsson (2016) for discussion of a moral analogoue of the principal principle that would rule out such behaviour as irrational.

tation theorem for moral uncertainty, in which case the Dutch book argument loses its standing as an independent argument for probabilism. Moreover, as I argued in section 3.4.1, this means that we can't get a fully general argument for probabilism with respect to moral uncertainty, because not all moral theories are comparable.

I therefore take it that a more promising line is to say that an agent may well have moral beliefs and be morally uncertain even if they believe that when it comes to deciding what one has all-things-considered most reason to do, prudence always trumps morality. Moreover, this combination of attitudes is rationally permissible, at least in the sense of rationality that Bayesians tend to be concerned with. This allows us to establish probabilism with respect to moral uncertainty for simple agents. As before, if we can then show that the assumption of linearity played no essential role in the argument, then we can take ourselves to have established for agents in general, and not just those who value money linearly.

Package Principle

The package principle does not seem to pose any further problems in the moral case. We have stipulated that simple agents value money linearly (and do not value anything else), and hence the package principle cannot fail for reasons of non-linearity. And we can further stipulate that in all relevant cases, simple agents do not regard their own betting behaviour as having any evidential bearing. And as before, this assumption is not essential for establishing the conclusion.

Sure Loss?

The question of what it means for a loss to be sure becomes especially pertinent when the propositions on which we are betting are moral ones. According to a common strand of thinking in meta-ethics, any true moral claim is necessarily true, at least if we consider moral claims of a sufficiently general nature (references). For example, it is generally believed that if utilitarianism is true, then it is necessarily true, i.e. true in all (ordinary) possible worlds. But we clearly don't wish to conclude that anyone who assigns utilitarianism a credence of less than one is thereby irrational.

Mahtani (2015) proposes an account of what it is for a loss to be sure that is well-suited to our purposes. She begins by observing that in the case of outright belief, an agent's belief state is coherent if and only if the set of all claims she believes is logically consistent. In turn, a set of claims is logically consistent if and only if there is an *interpretation* under which those claims are all true. An interpretation assigns meanings to all the non-logical terms in the language. For example, consider the claim "All triangles are asparaguses," which we can render in first-order logic as $\forall x T(x) \rightarrow A(x)$. Here the non-logical terms are $T(x)$ and $A(x)$. On one interpretation, $T(x)$ means "x is a hippopotamus" and $A(x)$ means "x is a mammal." Consider now the set containing only the claim "All triangles are asparaguses." According to our interpretation, this claim

means that all hippopotami are mammals, which is true. Since this is the only claim in the set, our interpretation renders true all claims in the set, and that set is therefore logically consistent. By contrast, consider the set containing the claims "All triangles are asparaguses" and "It is not the case that all triangles are asparaguses." In this case, there is no interpretation under which all claims in the set are true, and that set is therefore logically inconsistent.

Mahtani then extends this idea to degrees of belief, as follows. Take some set of bets that the agent regards as fair or better, and vary the interpretation of the claims involved. We can now say that the agent faces a sure loss just in case she makes a loss under every interpretation. So, just like an agent's outright beliefs are inconsistent just in case there is no interpretation according to which they are all true, so too her degrees of belief are incoherent just in case betting in accordance with those degrees of belief would lead her to make a loss under every interpretation.

This account allows us to address the issue that arose from the possibility that some moral claims may be true as a matter of metaphysical necessity. Suppose for the sake of illustration that utilitarianism is necessarily true. And suppose we gloss utilitarianism as saying that one option is better than another just in case it leads to greater happiness: $\forall x \forall y B(x, y) \leftrightarrow GH(x, y)$. An agent who regards £0.50 as the fair price for a bet that pays £1 if utilitarianism is false and £0 if it is true is not thereby incoherent, because it is not the case that she would make a loss under every interpretation.

In conclusion, the upshot seems to be that, despite initial concerns, Dutch book arguments work more or less just as well in the moral case. Therefore, insofar as you accept probabilism for descriptive uncertainty on the basis of a Dutch book argument, you should also accept it for moral uncertainty. However, this assumes that agents may rationally evaluate bets in a way that is contrary to her moral convictions. If you reject this assumption, you will find the Dutch book argument less compelling.

3.6 ACCURACY ARGUMENTS

As many have noted, representation theorem arguments (at least those of the decision-theoretic variety) and Dutch Book arguments (at least before given a depragmatized reading) both have a practical flavour. Decision-theoretic representation theorem arguments appeal to rational constraints on preference (or on evaluative judgments), and Dutch Book arguments appeal to undesirable financial consequences. Yet the conclusion both arguments seek to establish—probabilism—is an epistemic one, and it would therefore be desirable to derive it on epistemic grounds alone. Accuracy arguments for probabilism promise to do just this. They proceed from the assumption that the goal of our beliefs is to be accurate, and show that, in a certain specific sense, credences that satisfy probabilism are guaranteed to be more accurate than credences that violate probabilism, no matter what state of the world is actual. That is, if your credence function violates probabilism, there will always be some

specific credence function which satisfies probabilism that will be more accurate no matter what. Therefore, concludes the argument, you are rationally required to have probabilistic credences.[23]

There are two main components. The first is the claim that the goal of belief is accuracy, i.e. that accuracy is the only fundamental epistemic value. Call this claim *credal veritism*. The second is the claim that this notion of accuracy should be measured in a particular way. In the case of full belief it is quite clear what accuracy amounts to: a belief is accurate if it is true and inaccurate if it is false. But for degrees of belief it's not quite so straightforward. However, a natural starting point is the thought that accuracy somehow measures closeness to truth. A credence is more accurate the closer to the truth it is. Hence if A is true, then a higher credence in A should receive a higher accuracy score than a lower one. Conversely, if A is false, then a lower credence in A should receive a higher accuracy score than a lower one. Any acceptable way of measuring accuracy should have this property.[24] A credence function is a function $C : \mathcal{F} \mapsto \mathbb{R}$. Let us assume that credences take values in the unit interval, so that a credence of 0 in A corresponds to being certain that A is false, and a credence of 1 corresponds to being certain that it is true. This means that if A is false then 0 is the most accurate credence, and if A is true then 1 is the most accurate credence. Hence for a given world ω, the most accurate credence function is the one that assigns 0 to every claim that is false at ω and assigns 1 to every claim that is true at ω. Let us call this the credence function V_ω that is *vindicated* at ω. In order to measure the accuracy of some given credence function at world ω, we measure how close it is to V_ω.

Suppose we wish to compare two credence functions C_1 and C_2 (assumed to be numerical, but not assumed to be probabilistic) and determine which one is more accurate. To do so, we need to know which world is actual. If it will in fact rain tomorrow, then a higher credence in this claim will be more accurate than a lower one. But if it won't, then the lower credence will be more accurate. Hence accuracy can only be measured relative to a world. However, suppose we discover that C_1 is more accurate than C_2 relative to every world. That is, no matter what the actual facts are, C_1 is guaranteed to be more accurate than C_2. Clearly, C_1 is superior: an agent with credence function C_2 would be sure to become more accurate if she adopted C_1, and she could realise this *a priori*. If C_1 is at least as accurate as C_2 in every world, we say that it *weakly dominates* C_2. If C_1 is strictly more accurate than C_2 in every world, we say that C_1 *strongly dominates* C_2.

Let us also introduce the notion of expected accuracy. The expected accuracy of a credence function is calculated not relative to a world, but relative to another credence function. That is, I can determine how accurate I expect someone else's credences to be, in light of my own credences. This way, I can compare

23 See Rosenkrantz (1981), Joyce (1998) and Pettigrew (2016a).

24 Goldman (2002: 58) proposes that "the highest degree of belief in a true proposition counts as the highest degree of 'veritistic value' [...] In general, a higher degree of belief in a truth counts as more veritistically valuable than a lower degree of belief in that truth."

other people's credences in terms of how accurate I expect them to be. But the notion of expected accuracy also allows me to determine how accurate I expect my own credences to be. And we can then ask whether I expect myself to be more accurate than other people. Suppose I discover that I expect another credence function to be more accurate than my own. This creates a certain instability in my beliefs: by my own lights, I would do better by adopting this other credence function instead. But credences, at least if they are to be rational, should not undermine themselves in this way; they should, so to speak, be self-recommending. Therefore, any acceptable accuracy measure should be such that all rational credence functions expect themselves to be more accurate, according to that accuracy measure, than every other credence function. Accuracy measures that meet this condition are said to be strictly proper. For example, a commonly used strictly proper accuracy measure is the *Brier score* (Brier 1950):

$$\mathcal{B}(C, \omega) = 1 - \sum_{A \in \mathcal{F}} \mid V_\omega(A) - C(A) \mid^2$$

However, there are many other strictly proper accuracy measures. Why should we pick the Brier score rather than another? Luckily we don't have to because we can appeal to the following result. According to *any* strictly proper accuracy measure, (1) every non-probabilistic credence function is strongly dominated by some probabilistic credence function, and (2) no probabilistic credence function is weakly dominated by any other credence function (Predd et al 2009). That is, for every non-probabilistic credence function there is a probabilistic one which is guaranteed to be strictly better, in terms of its accuracy, in every possible world. But if the credence function is probabilistic, there is no other credence function (probabilistic or not) which is even just as good, in terms of its accuracy, as the probabilistic credence function. Therefore, insofar as we agree that the goal of belief is accuracy, and that accuracy should be measured by a strictly proper accuracy measure, we should conclude that credences are rationally required to be probabilistic.

3.6.1 *Objections*

Objections to accuracy arguments come in two forms: we might take issue either with the claim that accuracy is the only goal of belief, or with the claim that only strictly proper accuracy measures are acceptable. We shall focus here on the first.

Why think that only accuracy matters? After all, epistemologists have been concerned with various other aspects of our beliefs as well, such as whether they hang together in a coherent way (BonJour, 1985), or whether they are proportioned to the evidence (Conee and Feldman, 2004). But if these other things are epistemically valuable, then an argument which assumes that only accuracy is of epistemic value cannot establish probabilism, because for all that has been said we cannot rule out the possibility that these other epistemic values mandate in favour of non-probabilistic credences. The standard response on behalf of the accuracy theorist is to argue that insofar as these other things are

indeed of epistemic value, their value is not intrinsic, but can in fact be derived from accuracy considerations. In the first case, Pettigrew (2016c) notes that if the sort of coherence the coherentist has in mind amounts to probabilistic coherence, then the accuracy argument for probabilism vindicates coherence as well.[25] In the second case, he suggests that a set of accuracy arguments for epistemic principles beyond probabilism can underwrite the idea that rational agents proportion their credences to their evidence. In particular, there are accuracy arguments for conditionalization (Greaves and Wallace 2005), for the principal principle (Pettigrew 2013), and for the principle of indifference (Pettigrew 2016b). Together, these principles seem to support the idea of proportioning one's credences to the evidence. However, as we saw in the previous chapter, some believe that this evidentialism entails that, for at least certain bodies of evidence, imprecise credences are rationally required, or at the very least rationally permissible. And some have argued that no reasonable accuracy measure can vindicate the rationality of imprecise credences.[26] I take it to be an open question to what extent accuracy-based epistemology is compatible with imprecise credences.

3.6.2 *Accuracy Arguments and Moral Uncertainty*

One immediate difficulty for adapting accuracy arguments to the case of moral uncertainty is the following. If we define dominance in terms of greater accuracy in every (ordinary) possible world, and if we take the correct moral theory to be necessarily correct, then it follows that the credence function which assigns 1 to the correct moral theory and 0 to every other moral theory will be more accurate than every other credence function (whether probabilistic or not) in every possible world. Hence this credence function accuracy-dominates every other credence function. Therefore, it seems as though an accuracy argument for probabilism would yield the much stronger (and undesirable) conclusion that we always ought to adopt this unique maximally accurate credence function. But a natural response to this is to say that even if these other moral theories are not metaphysically possible, they are nevertheless epistemically possible, and this should be enough to get an accuracy argument going. After all, accuracy arguments probabilism do not seek to establish that we should have credence one in all metaphysical necessities, whether those metaphysical necessities are ethical or belong to some other domain. We can adapt Mahtani's account of sure losses to give an account of what it is for one credence function to accuracy-dominate another: for two credence functions defined over the same domain, we say that one accuracy-dominates another just in case it is more accurate than the other under every interpretation of the claims in the domain. This means that all of the notions that the accuracy argument appeals to are well-defined, and hence that we can formulate an accuracy argument for probabilism with respect to moral uncertainty. Let us now again consider the two types of objections to accuracy arguments, this time in the context of

25 Of course, there are other ways of measuring the coherence of a credence function, not all of which allow for coherence to be justified on accuracy grounds. See Olsson (2005).

26 See e.g. Seidenfeld et al (2012) and Schoenfield (2017a).

moral uncertainty.

First, do we have any particular reason to think that credal veritism fails when the beliefs in question concern moral claims? On the face of it, it seems that we don't. In order for credal veritism to fail, there would have to be properties of belief states that are epistemically laudable in the case of moral belief but not in the case of descriptive belief. If the credal veritist can give a plausible account of coherence and evidential proportionalism in the descriptive case, she can presumably also do so in the moral case. So in order to show that credal veritism fails, we would not only have to show that there are in fact these additional properties in the moral case, but also that —unlike coherence and evidential proportionalism—those properties cannot be accounted for in veritist terms. This strikes me as unlikely. Of course, it could also be that credal veritism fails in the descriptive case and *ipso facto* in the moral case. If so, the accuracy argument for probabilism with respect to moral uncertainty would fail, but not for reasons that are particular to morality.

Second, do we have any particular reason to think that the conditions on a reasonable accuracy measure are any different in the moral case than in the descriptive case? It seems not: whatever it is that make a credence function over a descriptive algebra accurate should also be what makes a credence function over a moral algebra accurate. Therefore, it seems that if you find accuracy arguments for probabilism persuasive in the descriptive case, you should do so in the moral case as well. Of course, I have not said anything here about how accuracy arguments for probabilism with respect to moral uncertainty fares on an expressivist understanding of moral language (see Staffel forthcoming), so it may well turn out that the very cognitive-sounding language of accuracy cannot be given a plausible expressivist reading, or that if it can, then the resulting expressivist notion of accuracy cannot underwrite an argument for probabilism. I leave these important questions for future investigation.

3.7 CONCLUSION

We have surveyed four types of argument for probabilism: decision-theoretic and epistemic representation theorem arguments, Dutch book arguments, and accuracy arguments. Of the four, the decision-theoretic representation theorem arguments was found most clearly wanting. Such arguments give us a utility function as well as a probability function, thereby forcing us to accept strong comparativism (i.e. the view that intertheoretic comparisons of value are always possible), at least if we are to obtain a fully general argument for probabilism with respect to moral uncertainty. Moreover, if we reject strong comparativism (i.e. we believe that there is at least some pair of incomparable theories), then we cannot appeal to coherent extendability here.

By starting from an epistemic rather than an evaluative notion, epistemic representation theorem arguments were able to sidestep the main issue with decision-theoretic representation theorem arguments. However, I conjectured that rational failures of completeness may be more frequent in the moral case

than in the descriptive case. If so, the kind of probabilism that the epistemic representation theorem arguments gives rise to will be more imprecise in the former than in the latter.

I argued that a depragmatized understanding of Dutch book arguments can allow them to vindicate probabilism with respect to moral uncertainty. However, this argument rests on the assumption that agents may rationally evaluate bets in a way that is contrary to their moral convictions.

Finally, I argued that if you think the accuracy argument for probabilism is sound in the case of descriptive uncertainty, you should also think it is sound in the case of moral uncertainty.

Where does this leave us?

The first thing to note is that we have not explored the possibility of giving these arguments a non-cognitivist reading in any detail. Therefore, any conclusions only hold conditional on congitivism.

Second, what conclusions you draw based on my examination will depend on how you think the arguments fare in the case of descriptive uncertainty. For example, Christensen (2004:124) writes that "If DBAs are the best-known ways of supporting probabilism, Representation Theorem Arguments (RTAs) are perhaps taken most seriously by committed probabilists." What Christensen has in mind are decision-theoretic representation theorem arguments in particular. It is therefore interesting to note that these are the ones that seem to fare worst in the case of moral uncertainty. Consequently, someone who believes that, of the ones considered, the decision-theoretic representation theorem argument is the only sound argument for probabilism with respect to descriptive uncertainty may reasonably reject probabilism with respect to moral uncertainty.

There may be other interactions between the arguments as well. For example, some have claimed that accuracy arguments are incompatible with imprecise credences (Schoenfield 2017). If I am right that epistemic representation theorem arguments more plausibly support imprecise probabilism than precise probabilism, it would seem to follow that we cannot accept both arguments: one makes imprecision rationally permissible whereas the other does not. But perhaps such an inference would be too quick, for there are some proposed accuracy measures that do vindicate imprecise credences (Mayo-Wilson and Wheeler 2016, Konek forthcoming).

At the very least, I take this chapter to have shown that probabilism with respect to moral uncertainty is sufficiently plausible to merit further investigation.

4

BAYESIAN MORAL EPISTEMOLOGY AND REFLECTIVE EQUILIBRIUM

4.1 INTRODUCTION

In the previous chapter, we examined various arguments for probabilism with respect to moral uncertainty and tentatively concluded that they succeed, with two caveats. First, we did not sufficiently explore the possibility of giving these arguments an expressivist reading, and consequently any tentative conclusions only hold conditional on cognitivism. Second, I suggested that failures of completeness may be more common in the case of moral uncertainty. Strictly speaking then, we have at best established imprecise probabilism. However, going forward I will mostly only consider agents with precise credences. Although precision is not rationally required, it is often rationally permitted, and it is therefore instructive to study the rationality constraints on agents with precise credences.

Once we have established something like probabilism, a natural next step is to ask whether we can also establish conditionalization. We would then have obtained both of the two core norms of traditional Bayesianism for moral uncertainty. And many arguments for conditionalization mirror those for probabilism. For example, there are both accuracy arguments (Greaves and Wallace 2005, Briggs and Pettigrew forthcoming) and Dutch book arguments (Teller 1973) for conditionalization.[1] Therefore, insofar as we accepted probabilism on the basis of such arguments, it seems we have reason to accept conditionalization as well. This means that all of the conceptual tools of Bayesianism become available for us to use. We can ask whether there are any constraints (beyond the probability axioms) on how we may assign prior probability to moral claims. We can ask whether any reasonable prior probability function is such that it satisfies the conditions of merging of opinion results. We can ask whether there are cases where Jeffrey conditionalization is more appropriate than standard conditionalization. We can ask how our degrees of belief in moral claims relate to our full beliefs in moral claims. And we can ask whether one can plug the probabilistic credences into a theory of decision-making under moral uncertainty.

However, even if the arguments for conditionalization succeed, it is still possible that we end up with a moral epistemology that looks rather different from

1 Although see Mahtani (2012) and Schoenfield (2017b).

traditional Bayesian epistemology. This is because the rule of conditionalization only tells agents to conditionalize if they have learned some proposition with certainty. It doesn't say anything about how agents should behave if their learning experience does not take this form. Therefore, if learning experiences that cannot be captured using conditionalization (or Jeffrey conditionalization) are more widespread in the moral case than in the descriptive case, the resulting epistemological picture will be different. In fact, I will argue that something like this is the case. I will argue that in order to model certain aspects of moral epistemology, we need to avail ourselves to the possibility of *awareness growth*. When an agent's awareness grows, she becomes aware of possibilities she had not previously entertained, and therefore not previously assigned any probability (or taken any other doxastic attitude towards). Traditional Bayesianism cannot capture this, because it assumes that the σ-algebra $\mathcal{F}$ over which the agent's probability function is defined remains fixed over time.

Although the phenomenon of awareness growth is familiar from the descriptive domain, I will suggest that it may play an epistemologically more central role in the moral domain. In particular, I will argue that if we wish to make our Bayesian moral epistemology compatible with the method of reflective equilibrium, we must make room for awareness growth. Since I take it that reflective equilibrium, in one form or another, has the status of standard method in moral epistemology, it would count against Bayesian moral epistemology if the two turned out to be incompatible. But happily, this turns out not to be the case. Or so I will argue.

This is our third variation then: what do the dynamics of credence look like when the learning experience does not take the form of probabilities shifting across some partition, but instead consists in the agent becoming cognizant of new possibilities? First, however, we will briefly look at some cases in which conditionalization does seem to get moral epistemology right (section 4.2). After that, I discuss some cases of moral learning where neither conditionalization nor Jeffrey conditionalization seem applicable, and argue that in order to account for such cases, we must make room for awareness growth (section 4.3). I then introduce the method of reflective equilibrium, and argue that this method also calls for awareness growth (section 4.4). Finally, I present a formal account of reflective equilibrium (sections 4.5-4.6).

4.2 BAYESIAN MORAL EPISTEMOLOGY

As you will recall, this is how we formulated the rule of conditionalization:

CONDITIONALIZATION If a rational agent with probability function $P(\cdot)$ learns proposition E with certainty and nothing else, her new probability function is given as $P_E(\cdot) = P(\cdot \mid E)$

As the formulation makes clear, the rule only applies when there is some proposition that the agent learns with certainty. You might wonder how often this condition is met when it comes to our moral beliefs. In the ordinary

empirical case we can make observations and conditionalize on those observations. Although it's certainly not unproblematic to say that we should become certain of our empirical observations, if anything the moral case looks even more dubious. What observations can confirm or disconfirm moral theories? One possibility is that they are the same kind of empirical observations as those that confirm or disconfirm scientific theories. If an agent believes that some observation is more likely on one moral theory than on another, then making that observation confirms the former and disconfirms the latter. And if we are happy to conditionalize on ordinary empirical observations in a scientific context, we should be equally happy to do so in a moral context. So in principle there is no problem here. But we might still wonder which empirical observations may *reasonably* be taken to count in favour of some given moral theory. Here, I survey a few possibilities. My discussions will be brief, and I certainly don't take them to establish that conditionalization is absolutely the right way to model all of these cases. However, I do take them to at the very least make plausible the idea that there is some interesting role for conditionalization to play in a probabilistic moral epistemology.

Testimony

A paradigmatic example is moral testimony. At least on occasion, it seems reasonable to treat the moral views of other people as evidence, especially if you judge them to be trustworthy, reflective, and wise (or to possess whatever set of abilities you take to be conducive to having accurate moral beliefs).[2] In a case of testimony, the proposition that is learnt with certainty takes the general form "Person A reports that N" but the type of moral claim being reported may vary. For example, they might report that some option is impermissible, or that one option is better than another, or that some aspect of an option is a morally relevant consideration, or that some general moral theory is correct, and so on. For any such type, they may either assert the claim itself, or assert that they have some particular credence in it.[3]

Given that some proposition is learnt with certainty, standard conditionalization seems appropriate. If the agent takes the person to be reliable, the result of conditionalizing will be to increase her credence in the claim being reported. Furthermore, she will also increase her credence in other moral claims that entail the reported claim, and decrease her credence in claims that entail its negation. For example, if they report that it's impermissible to switch in a trolley case, she will increase her credence in Kantian deontology and decrease her credence in utilitarianism. If they report that equality of outcome is a morally

2 Many philosophers have argued that there is something epistemically defective about relying on testimony when the subject matter is morality. The same philosophers are typically happy to allow that reliance on testimony can be perfectly acceptable in other domains, and hence they argue that there is something distinctive about morality which makes testimony inappropriate as a basis of belief. However, most discussions of moral testimony assume that belief is an all-or-nothing matter, as opposed to the graded picture we are working with here. See Hills (2013) for an overview of existing work on moral testimony.

3 Of course, they may also report having other types of doxastic attitudes towards the claim, such as being fairly confident, or being agnostic, etc. But for simplicity I will restrict attention to assertions and credence reports.

relevant consideration, she will increase her credence in *ex post* egalitarianism and decrease her credence in Nozickian libertarianism. And so forth.[4]

In the case where the person also reports their credence in the moral claim, we might consider a moral deference principle, analogous to other expert principles such as the principal principle or the reflection principle. To treat someone like a moral expert in this sense is to satisfy something like the following condition: $P(N \mid P_A(N) = x) = x$, where $P_A(N)$ is person A's credence in N. Formally speaking, the fact that the domain of inquiry concerns moral matters does not pose any distinctive problems for the formulation of expert principles. Nevertheless, you might wonder how one could be justified in treating someone as a moral expert in this sense. When deciding to treat, say, a meteorologist as a weather expert, one is able to consult their track record and justify one's decision on that basis. But when the purported expertise concerns morality, it seems that no such track record is available (or if it does exist, it will be much smaller). Of course, and as we have seen, you might regard a person's testimony as evidence without treating them as an expert in the relevant sense. In the case where someone simply asserts a moral claim, you treat that assertion as positive evidence just in case $P(N \mid A \text{ asserts } N) > P(N)$. So even if you think that one should never (or only very rarely) treat someone as a moral expert in the sense of deferring to their credences, it is clear that moral testimony can still play a role.

Emotions as Evidence

Another possible source of morally relevant empirical observations is the agent's own reactions. For example, if she takes feelings of guilt to reliably indicate wrongness, then whenever she comes to feel guilty about something she's done, she should increase her credence that what she did was impermissible. Of course, it is crucial here that she actually does believe that feelings of guilt track wrongness. If she instead believes a debunking explanation according to which feelings of guilt simply serve to make one an obedient member of society (and further believes that this has nothing to do with morality), then she should not treat feelings of guilt as evidence in this manner.

For agents who do take feelings of guilt to be relevant, it seems appropriate to model such cases using conditionalization. The agent notices that she is feeling guilty about something she's done, and conditionalizes on this fact. As a result, she will increase her credence in moral claims according to which she did something impermissible and decrease her credence in those according to which she did not. All of this is of course assuming that she was at least somewhat surprised by her feelings of guilt: if she knew that she would feel guilty but carried on anyway, then those feelings do not provide any new information. Of course, guilt is just one salient example here: there may be other feelings that agents take to carry some moral information, and these can be treated in the same way.

4 Enoch (2014) argues that deferring to experts is sometimes the only morally acceptable way of responding to moral uncertainty.

Track Record

In the sciences, theories gain support in part by their track record. All else equal, a scientific theory that makes more accurate predictions deserves higher credence than one which makes fewer. Might the same idea apply in the moral case? One obvious difference is that in the scientific case, there is much wider interpersonal agreement on what the data points are, and therefore on which predictions were accurate and which were not. By contrast, there is very little agreement on what the data points are in the moral case. But even if there is little agreement, there might still be some, and that is all we need to get the idea going.

Consider for example an early utilitarian like Bentham. On the basis of his utilitarianism, he argued in favour of freedom of expression, equal rights for women, the decriminalising of homosexual acts, and animal rights; and against the death penalty and physical punishment. In many instances, these views flew in the face of the prevailing wisdom of his time. Nonetheless, there is today widespread agreement that he got these things essentially right. Of course, the agreement is still far from unanimous, and we should be wary of treating contemporary opinion as a perfect guide to truth. But it is still striking how much Bentham seems to have gotten right. Contrast this with Kant, who opposed homosexuality, claimed that husbands possess their wives in much the same way that they possess material objects, argued that masturbation is a grave sin, and held that it's permissible to kill children born out of wedlock.[5] In these cases, there is today widespread agreement that Kant got these things wrong. Again, that agreement is far from unanimous, and certainly not as clear-cut as the agreement on what the data points are in the scientific case. Nevertheless, to the extent that we do take these to be data points, we can regard them as counting in favour of utilitarianism and counting against Kant's moral philosophy. Many of Bentham's 'predictions' have been borne out, whereas many of Kant's have not.

Of course, the above example is merely intended to illustrate the possibility in principle of testing moral theories in a way analogous to the testing of scientific theories. It would take a lot more careful examination to justify the particular inference that the historical track record supports utilitarianism over Kantian ethics. Perhaps the difference between Bentham and Kant does not tell us anything about their moral philosophies, but merely something about their temperaments, with Bentham more disposed to think through the implications of his view in a detached manner, and Kant more sensitive to the spirit of his time. Furthermore, a complete assessment would need to look at their entire bodies of work, to make sure that we're not cherry-picking our examples. In particular, utilitarianism does of course have many other implications that sound unpalatable to contemporary ears.

In conclusion, I hope to have made plausible the claim that conditionalization has a role to play in moral learning. We sometimes revise our moral credences

5 Schwitzgebel (2010).

by updating on the testimony of someone else. If we take guilt and other moral feelings to indicate wrongness, then such feelings will also lead us to revise our moral credences. And if we think there are fairly uncontroversial 'data points' that any theory should account for, we can use those to assess theories. However, many types of moral learning are left out of this picture. Some might be dealt with by generalizing conditionalization to Jeffrey conditionalization to allow for uncertain learning. Conflicting testimony or mixed emotions are perhaps naturally modelled as cases of Jeffrey conditionalization. But I will argue that many types of moral learning are left out even by this more general picture. Some such cases are better thought of as cases of awareness growth.

4.3 AWARENESS GROWTH

Suppose that you come across the following hypothetical scenario for the first time:

> You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you, "Look, we're sorry the Society of Music Lovers did this to you—we would never have permitted it if we had known. But still, they did it, and the violinist now is plugged into you. To unplug you would be to kill him. But never mind, it's only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from you." Is it morally incumbent on you to accede to this situation? (Thomson 1971:49)

Given that you had never come across the scenario before, it seems as though it would be inaccurate to say that you assigned some particular credence to the claim that it would be morally required for you to accept the situation. But the trouble is not with precision: imprecise credences would not provide a more plausible model, and neither would suspension of judgment. It seems more natural to say that, prior to encountering the scenario, you took no doxastic attitude whatsoever towards the relevant claim, because you were unable to entertain it in the first place.

A fundamental feature of the traditional Bayesian picture is that the σ-algebra $\mathcal{F}$ over which the probability function is defined remains fixed throughout the agent's epistemic life. She goes about her day revising her credences (via conditionalization or Jeffrey conditionalization) in light of the information she receives, perhaps also making some decisions along the way. But her fundamental conception of the space of possibilities never changes: all that ever happens, epistemically speaking, is that her probability assignment shifts around

between the various propositions in $\mathcal{F}$. When an agent undergoes *awareness growth*, she becomes able to conceive of possibilities she was previously unable to entertain. For example, suppose that the agent has never before come across the idea that we might one day be able to create artificial intelligence that vastly outperforms humans on a wide range of tasks. Given that she was previously unable to entertain this proposition, it was not an element of her σ-algebra $\mathcal{F}$, and therefore she did not previously assign it any probability.

Clearly, this sort of thing happens to us all the time. We don't have a maximally fine-grained understanding of the entire space of possibilities; instead that understanding is constantly evolving as new possibilities become salient to us. So if a normative Bayesian theory is to apply to agents like ourselves, it must allow for the possibility of awareness growth. Let us therefore see what a Bayesian account of awareness growth might look like.[6]

4.3.1 *Formal Account of Awareness Growth*

There are two salient ways in which an agent's awareness may grow. In cases of *expansion*, new possibilities are added to the sample space Ω. In cases of *refinement*, the sample space remains the same, but the propositions in $\mathcal{F}$ are individuated more finely.[7] The distinction is perhaps best appreciated by way of example.

EXAMPLE 1 You are thinking about tomorrow's weather. At first you take the possibilities to be a clear or rainy sky, and a warm or cold temperature. But then you realise that if it's a clear, warm day, it could be either windy or calm, and similarly for the other possibilities.

EXAMPLE 2 You are again thinking about tomorrow's weather, and your initial conception of the relevant possibilities is the same as before. But then you realise that, in addition to clear and rainy, the sky could also be snowy.

	Clear	*Rainy*
Warm	WC	WR
Cold	CC	CR

Table 3: *Initial Awareness Context*

In both cases, your initial awareness context is given by table 1. In example 1, you *refine* it to arrive at table 2. In example 2, you *expand* it to arrive at table 3. Formally, what happens is that the belief state changes from $\langle \Omega, \mathcal{F}, P \rangle$ to $\langle \Omega^+, \mathcal{F}^+, P^+ \rangle$, with $\Omega \subseteq \Omega^+$ and $\mathcal{F} \subset \mathcal{F}^+$. In the case of refinement, $\Omega^+ = \Omega$ and $\mathcal{F} \subset \mathcal{F}^+$. In the case of expansion, $\Omega^+ = \Omega \cup X$, and $X \in \mathcal{F}^+$.

6 The phenomenon of awareness growth is closely related to the problem of new theories for Bayesian confirmation theory (Glymour 1980).

7 I take the terminology of refinements and expansions from Bradley (2017: 258).

	Clear	*Rainy*
Warm & Windy	WWC	WWR
Warm & Calm	WCC	WCR
Cold & Windy	CWC	CWR
Cold & Calm	CCC	CCR

Table 4: *Refinement*

	Clear	*Rainy*	*Snowy*
Warm	WC	WR	WS
Cold	CC	CR	CS

Table 5: *Expansion*

You might think that all instances of awareness growth can be captured as refinements by requiring that the agent's awareness context $\mathcal{F}$ always include a "catch-all" proposition (Shimony 1970). Intuitively, the catch-all proposition says something like "none of the above." We can then model example 2 as a case of refining the catch-all proposition into a new proposition, *Snowy*, and a new catch-all proposition. And more generally, we can represent the initial awareness context as in table 6.

	Clear	*Rainy*	*??*
Warm	WC	WR	W??
Cold	CC	CR	C??
?	?C	?R	???

Table 6: *Using a "Catch-All" Proposition*

Here we have two catch-all propositions: *?* is the event that the temperature is neither warm nor cold, and *??* is the event that the sky is neither clear nor rainy. How should the agent assign credence among these propositions? In particular, how should she assign credence to propositions involving either of the two catch-alls? Some object to the approach of using a catch-all precisely on the grounds that it requires agents to assign probability to a proposition that they do not understand.[8] Although we often do recognise, and perhaps often should recognise, that there may be more possibilities than we have conceived of, this does not straightforwardly translate into a requirement to assign a precise numerical probability to these unimagined possibilities.

We have been discussing awareness growth in probabilistic terms, but given that the basic definitions formulated in terms of conditions on σ-algebras, they are just as readily applicable to categorical and relational belief. In the categorical case the belief state changes from $\langle \Omega, \mathcal{F}, \mathcal{B} \rangle$ to $\langle \Omega^+, \mathcal{F}^+, \mathcal{B}^+ \rangle$, and in the relational case it changes from $\langle \Omega, \mathcal{F}, \trianglerighteq \rangle$ to $\langle \Omega^+, \mathcal{F}^+, \trianglerighteq^+ \rangle$. In the following, however, we shall mostly be concerned with awareness growth for agents with probabilistic beliefs.[9]

4.3.2 *Rationality Constraints on Awareness Growth*

After an agent has undergone awareness growth, she must adopt a new probability function P^+ that assigns credence to the propositions in her new σ-

8 For example, Stefánsson and Steele (manuscript) voice this concern.

9 See Schipper (2014) for an overview of work on awareness growth.

algebra $\mathcal{F}^+$. How should she do this? Traditional Bayesianism is silent on the matter, because it assumes that the σ-algebra is fixed. Some have proposed ways of extending the Bayesian framework. A common theme among such proposals is that belief change in the face of awareness growth should be conservative in a way that is similar to ordinary updating. Roughly speaking, Bayesian updating is conservative in the sense that it tells agents to make the most minimal change in their beliefs necessary to bring those beliefs in line with what has been learnt. For example, Karni and Vierø (2013) propose the following constraint, which they dub *reverse Bayesianism*:

REVERSE BAYESIANISM For any $A, B \in \mathcal{F} \cap \mathcal{F}^+$:

$$\frac{P(A)}{P(B)} = \frac{P^+(A)}{P^+(B)}.$$

Reverse Bayesianism is conservative in the sense that it whenever an agent's awareness grows, she should preserve the probability ratios between all propositions that she entertained prior to the growth in awareness. In example 1, this means that

$$\frac{P(Warm)}{P(Cold)} = \frac{P^+(Warm \,\&\, Windy) + P^+(Warm \,\&\, Calm)}{P^+(Cold \,\&\, Windy) + P^+(Cold \,\&\, Calm)}$$

That is, learning that it can be either windy or calm should not affect the relative probability you assign to warm and cold. In example 2, reverse Bayesianism implies that learning that it could be snowy should not affect the relative probability you assign to clear and rainy:

$$\frac{P(Clear)}{P(Rainy)} = \frac{P^+(Clear)}{P^+(Rainy)}$$

However, reverse Bayesianism does not provide any constraints on what credence you may assign to *Snowy*. Nor does it constrain how you divide your credence between *Warm & Windy* and *Warm & Calm* in example 1. But reverse Bayesianism is arguably still too strong. Consider the following case (Stefánsson and Steele, manuscript):

> Suppose you happen to see your partner enter your best friend's house on an evening when your partner had told you she would have to work late. At that point, you become convinced that your partner and best friend are having an affair. You discuss your suspicion with another friend of yours, who points out that perhaps they were meeting to plan a surprise party to celebrate your upcoming birthday—a possibility that you had not even entertained. Becoming aware of this possible explanation for your partner's behaviour makes you doubt that she is having an affair.

Stefánsson and Steele suggest that it would be perfectly rational if, as a result of becoming aware of this new possibility, your credence that your partner is having an affair decreases proportionally more than your credence that she is overworked. But if so, reverse Bayesianism is violated. Bradley (2017:258) proposes a similar constraint, also based on the idea that agents should make the most minimal change necessary:

AWARENESS RIGIDITY For any $A \in \mathcal{F} \cap \mathcal{F}^+$, $P^+(A \mid \Omega) = P(A)$.

Reverse Bayesianism implies Awareness Rigidity, but not the other way around.[10] Awareness Rigidity and Reverse Bayesianism will give the same result in cases of refinement, but not in cases of expansions.[11] Given that the counterexample to Reverse Bayesianism was a case of expansion, Awareness Rigidity is not subject to the same counterexample, because it does not constrain credences as strongly.

This concludes our survey of rationality constraints on awareness growth. We now turn to the main area of application: reflective equilibrium.

4.4 REFLECTIVE EQUILIBRIUM

4.4.1 *Background*

If there is anything like a standard method in moral epistemology, it is arguably some form or other of the method of reflective equilibrium.[12] Rawls originally introduced the term in the context of his theory of justice as fairness, but the general framework has been applied much more broadly.[13] In what follows, I will therefore assume that all talk of reflective equilibrium concerns our moral beliefs in general, not just those that concern justice.

First, we imagine that there is some agent engaged in moral deliberation. We make certain requirements of this agent: she "is presumed to have the ability, the opportunity, and the desire to reach a correct decision (or at least, not the desire not to)" (Rawls 1999:42) Second, we assume that there are two types of moral claims: *particular judgments* and *theoretical principles*.

The particular judgments are in some sense *specific*. For example, they may concern particular acts, such as the judgment that it was morally wrong for David Cameron to call the Brexit referendum. Furthermore, they don't cite the reasons for their verdicts, e.g. the particular judgment does not explain why it was wrong to call the referendum. The theoretical principles, on the other hand, are *general*: they are applicable in a wide range of circumstances. Moreover, they are explanatory: they are able to cite reasons for their verdicts. In particular, they are able to explain the various particular judgments that the agent holds.

The agent engaged in moral deliberation is not a blank state: she begins deliberation having already made some moral judgments. We may impose constraints on which sets of moral judgments count as legitimate starting points. Rawls

10 To see the implication, simply replace all probabilities in the statement of reverse Bayesianism with probabilities conditional on Ω, and note that $P(A \mid \Omega) = P(A)$ and $P(B \mid \Omega) = P(B)$.

11 Note that in cases of refinement, $P^+(A \mid \Omega) = P^+(A)$, whereas in cases of expansion this will not generally be the case.

12 See Rawls (1971/1999), Scanlon (2003), Daniels (2018), and Cath (2016).

13 Rawls himself attributes this idea to Nelson Goodman's (1955) work on the justification of deductive and inductive rules of inference.

(1951) thought that the initial moral judgments all had to be particular; no theoretical principles allowed. Moreover, the particular judgments in question all had to concern actual (as opposed to hypothetical) cases. Rawls (1971:41) suggested that they were particular judgments "with their supporting reasons." And Rawls (1975) held that the agent's initial belief set was allowed to contain moral claims of any level of generality.

Throughout, Rawls emphasised that not just any set of moral judgments will do as a starting point. Those judgments also need to have a certain epistemic pedigree. In *A Theory of Justice*, he held that we can discard (i) judgments made with hesitation, (ii) judgments in which we have little confidence, (iii) judgments made when we are upset or frightened, and (iv) judgments made when we stand to gain one way or another. He refers to the judgments that remain after these have been discarded as the agent's *considered judgments*.

In the simplest case, we can think of the search for reflective equilibrium as follows. The agent starts out with only particular judgments. In trying to systematise and account for these particular judgments, she considers various theoretical principles, perhaps finding some of them quite plausible. However, it is unlikely that there will initially be a perfect fit between her particular judgments and her theoretical principles. She may have to give up some of her initial particular judgments and come to accept new ones. And she may also have to replace some of her theoretical principles with others. Eventually though, working her way back and forth between revisions to her particular judgments and revisions to her theoretical principles, the hope is that she will reach reflective equilibrium. When she has reached reflective equilibrium, her beliefs are consistent, and her theoretical principles account for her particular judgments in a satisfactory manner:

> It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation (Rawls 1999:18).

Of course, the reason we are interested in reflective equilibrium is that we think that the beliefs held in reflective equilibrium are epistemically laudable in some way: that they are justified, or that they are more likely to be correct than beliefs acquired through some other method, or something else along these lines. But some have worried that, in giving comparatively large weight to the agent's initial judgments, reflective equilibrium is an inherently conservative method, in the sense that any equilibrium the agent reaches is unlikely to deviate too far from her starting assumptions. But if so, it seems we may have reason to think that the method of reflective equilibrium may not be so trusthworthy after all. For if the result never deviates to far from the starting assumptions, we will not be able to resolve disagreement between agents with different starting assumptions.

Partly in response to this, Rawls (1975) distinguished between *narrow* and *wide* reflective equilibrium. One may reach reflective equilibrium in this narrow

sense rather easily, perhaps just by making a few minor adjustments to one's moral beliefs. But the philosophically more interesting notion of reflective equilibrium—wide reflective equilibrium—requires more than this. It requires that one's moral beliefs have withstood various kinds of scrutiny. This can be specified in various ways, and in principle there are no restrictions. It might require that one has given all other comprehensive moral theories on offer a serious consideration. It might require that the agent inform herself of various empirical facts (such as those of the social sciences) that may be relevant. Or it might require that the agent considers some other facts that have bearing on whether or not the beliefs held in reflective equilibrium are justified. Rawls (1975:8) says that wide reflective equilibrium is found after we have

> had an opportunity to consider other plausible conceptions and assess their supporting grounds. Taking this process to the limit, one seeks the conception, or plurality of conceptions, that would survive the rational consideration of all feasible conceptions and all reasonable arguments for them.

4.4.2 *Towards a Formal Theory*

Now that we have a rough idea of what reflective equilibrium looks like, we can start thinking about how to formalise it in a way that makes it amenable to a Bayesian treatment. Before we do so, let me make a couple of clarifications about the notion of reflective equilibrium I shall be proposing. First, we will not concern ourselves with how some initial judgments are discarded due to lack of epistemic merit. The initial judgments are given exogenously. Second, the notion of reflective equilibrium we shall formulate corresponds more closely to narrow than to wide. We will not be modelling the process by which agents consider all reasonable arguments for all feasible principles. On the other hand, if the critics are right that narrow reflective equilibrium is conservative in the sense that the results are heavily biased towards the initial judgments, then our notion of reflective equilibrium is not a narrow one either, because it may turn out that in reflective equilibrium an agent rejects all of her initial judgments. With those clarifications in mind, let us now examine what a formal account of reflective equilibrium might look like.

Consider the following highly stylised example. Suppose an agent initially judges that ordering steak for lunch yesterday was impermissible, and that saving the child drowning in the pond would be obligatory. She then reflects on what these two judgments have in common. Perhaps she realises that, of the available options, saving the child would maximise total happiness, whereas ordering steak would not. Insofar as she finds this a plausible explanation of her initial judgments, she may come to believe the theoretical principle which says that an option is permissible if and only if it maximises total happiness. If she does, she has reached reflective equilibrium: she has endorsed a theoretical principle that accounts for her particular judgments in a satisfactory manner.

The example is formulated in terms of categorical belief, even though it is reflective equilibrium for probabilistic belief that we are ultimately concerned with. However, much of what I will have to say applies equally well to each of the three models of belief we have considered (i.e. categorical, relational, and quantitative). In particular, the formal accounts I give of particular judgments and theoretical principles can be used in all of these models. And the same is true of the account of the process of reflective equilibrium I give in terms of awareness growth. Given that categorical belief is in many ways simpler than probabilistic belief, we shall initially formulate things in terms of categorical belief, only later moving on to consider Bayesian reflective equilibrium.

With that in mind, what lessons can we draw from the example? Let's begin with particular judgments. First, these judgments concern specific situations: the agent's lunch decision yesterday, and Singer's hypothetical scenario. They don't say anything about situations other than these. In particular, the judgment that ordering steak was impermissible does not imply anything at all about whether saving the child is obligatory or not. Second, particular judgments don't say anything about why these verdicts hold. This means that in principle, various different theoretical principles could account for the same set of particular judgments.

Next, theoretical principles. In contrast with particular judgments, theoretical principles are general: they concern a wide range of cases. Second, theoretical principles can both entail and explain particular judgments.

In the example, the agent initially only has beliefs about particular actions: she believes that ordering steak was impermissible and that saving the child would be obligatory, but she neither believes nor disbelieves that an option is permissible if and only if it maximises total happiness. You might think this means that she suspends judgment with respect to this theoretical principle, but that doesn't seem to get things right either. Suspension of judgment implies that she has considered the principle and is unable to make up her mind, but this does not appear to be a faithful characterisation of her state of mind. Instead, it seems to me that a more natural description of what's going on is that she has not even considered the possibility. Therefore, we need to allow for awareness growth. The agent starts out initially only aware of particular judgments, and then she formulates theoretical principles through awareness growth.

Perhaps this is the right thing to say about this toy example, but what about more sophisticated reasoners like ourselves? When we engage in reflective equilibrium reasoning, it's not as though we begin from a state of complete ignorance about general moral principles. We are familiar with various such principles, and don't have to formulate them all from scratch. But while this is surely right, it also does happen, at least on occasion, that we come across new principles that we hadn't considered before. Moreover, for all of the familiar moral theories, there are various ways of making them more precise. Even a comparatively simple moral theory such as hedonistic utilitarianism

can be made more precise in various ways. What exactly counts as pleasure? How should it be measured? Answering these questions and others gives us different versions of the theory. And therefore, we will still need to appeal to awareness growth when giving an account of reflective equilibrium for more sophisticated reasoners. However, to keep things simple, I will first focus on agents who are initially only aware of particular judgments and have to formulate all theoretical principles through awareness growth.

Moreover, there is a further role for awareness growth in the process of reaching reflective equilibrium. Just like we may become aware of new theoretical principles, we may also become aware of new particular judgments. Sometimes we are confronted with entirely new situations, whether in real life or in a hypothetical scenario presented to us, and we may then form new particular judgments about theses cases. If we discover that these new judgments are inconsistent with the theoretical principle we have tentatively settled on, we shall have to revise that theoretical principle. Indeed, this can be a way of testing a proposed theoretical principle.

In our toy example, the theoretical principle was said to account for the particular judgments. How shall we unpack this notion? A natural suggestion is that a theoretical principle accounts for a set of particular judgments in virtue of both entailing and explaining those judgments. This also gives us a natural way of saying what it takes for an agent to reach reflective equilibrium: she has reached reflective equilibrium just in case her belief set consists of some particular judgments together with a theoretical principle that entails and explains all of those judgments.

If, in equilibrium, a theoretical principle is required to entail all of an agent's particular judgments, it follows that an agent in equilibrium cannot simultaneously endorse multiple theoretical principles. You might think that this fits poorly with how theoretical principles are usually conceived of in the reflective equilibrium framework. For example, an agent may without inconsistency endorse both a principle of utility, according to which greater utility is better (all else equal), and a principle of equality, according to which greater equality is better (all else equal). In part, I take it that this is a merely verbal question: the agent who endorses both the principle of utility and the principle of equality can just as well be construed as believing a single principle, the principle of utility and equality, defined as the intersection of the other two principles. Whenever there is only utility at stake, the joint principle says that the greater utility option is best, and whenever there is only equality at stake, it says that the greater equality option is best. In cases where the two considerations pull in different directions, the principle is silent on which option is best.

The more substantive issue is whether a theoretical principle has to entail all of the particular judgments the agent endorses in equilibrium. The principle of utility and equality is silent on which option is best whenever the two considerations pull in different directions. But in some such cases the agent might still hold the particular judgment that the higher utility option is best, or that

the higher equality option is best. Perhaps she hasn't made up her mind about exactly how utility and equality should be traded off against one another.

But even if she hasn't settled on a precise conversion rate, the fact that she is able to make a judgment in some cases suggests that she could at least formulate a somewhat more precise principle that yields these verdicts as well. This more precise principle would better account for her judgments than the initial one. On the other hand, if there is some case in which she is genuinely unable to make up her mind as to which consideration should win out, then her theoretical principle should also remain silent. Any theoretical principle that gave a determinate verdict in such cases would misrepresent the beliefs she in fact holds.

Granted, it is rather demanding to require that, in equilibrium, an agent must endorse a theoretical principle that logically entails all of her particular judgments. But the fact that it is so demanding makes it a natural place to start. Clearly, mere consistency cannot be enough for reflective equilibrium: far too many far too strange theoretical principles will be consistent with a given set of particular judgments. So we should at least require that a theoretical principle must entail some of her particular judgments. But once we've conceded that it must entail some of them, the most natural place to stop is to say that it must entail all of them. On the face of it, any other stopping point would seem arbitrary. And even if it could be chosen in a non-arbitrary way, the resulting proposal would be a lot more complicated. So in the spirit of first getting a simple theory on the table, saving modifications for later, we should stick with the demanding requirement for now.

Another consideration in favour of the present proposal is that in some sense it represents an epistemic ideal. An agent who has formulated a theoretical principle that entails all of her particular judgments is in a better epistemic position than an agent who has formulated a theoretical principle that only entails some of them. The former principle explains everything, whereas the latter leaves some things unexplained. Again, we should therefore begin with the more idealised notion, saving possible modifications for later.

You might worry that the requirement that a theoretical principle entail all of the particular judgments introduces a risk of overfitting. The charge would be justified if her particular judgments were forever set in stone. But the requirement only comes into play provided that the agent has in fact reached reflective equilibrium. Initially, her particular judgments may be subject to all sorts of biases. But if she still holds on to them after having gone through multiple revisions in search of reflective equilibrium, it no longer makes sense to think of them as potential sources of noise. Instead, they are data points that an adequate theoretical principle should account for.

We can make some general observations about what this means in the particular formal framework that we shall be using. Given that we will be modelling both types of moral claims as sets of reasons structures, it follows that what it

is for a theoretical principle to entail a particular judgment is for the principle to be a subset of the judgment. Moreover, the notion of explanation should somehow be cashed out in terms of normatively relevant properties: what it is for a theoretical principle to explain a particular judgment is for it to say something about the normatively relevant properties in virtue of which that judgment holds. We will return to this in more detail later.

A closely related point is that theoretical principles are general: they apply in a wide range of cases. In our example, the principle applied both in the case of ordering steak and in the case of the drowning child. And as we have just seen, an adequate theoretical principle should be general enough to account for all of the particular judgments the agent holds in equilibrium. However, it should be still more general: it should be applicable even in cases the agent has not considered explicitly.

For example, the agent who formulates and endorses the theoretical principle according to which an option is permissible if and only if it maximises total happiness may then come to see that this principle has implications for cases beyond those she had initially considered. Perhaps she comes across Philippa Foot's case of the surgeon who cuts up a healthy person in order to distribute their organs and save the lives of five other people. When she does, she may form the particular judgment that it would be impermissible for the surgeon to do so. If so, she will have to give up her theoretical principle.

This points to another feature of theoretical principles that we should be able to capture in our framework: the agent may not realise all their implications and when she does, she may feel compelled to give up the principles in light of their implications.

How shall we think about this? One option is to see it as a failure of logical omniscience: the agent simply doesn't realise that her theoretical principle is implicitly inconsistent with her particular judgment about the case of the surgeon. But while this certainly happens from time to time, it doesn't seem like an accurate description of all cases. In order for this to be a case of implicit inconsistency, the agent must already have formed a particular judgment about the surgeon case. But if the agent has never considered the case before, she will not have formed a particular judgment about it. And just like it was implausible to say that the agent initially suspended judgment about the theoretical principle, so too it would be implausible to say that she initially suspended judgment about the surgeon case. Again, awareness growth provides a more natural framework for thinking about what's going on here.

Of course, the restriction to particular judgments and theoretical principles is somewhat artificial: there is no reason to think that all of our moral judgments can be neatly divided into these two categories. Ultimately, our concern is to give an account of reflective equilibrium for agents whose objects of belief are arbitrary sets of reasons structures. Nevertheless, by initially considering considering only these two categories of moral claims, we are able to focus our

attention on an important aspect of the methodology, namely the fact in order for an agent to be in reflective equilibrium, it is not enough that her beliefs be logically consistent: her beliefs about particular cases also need to be explained by a general principle.

Let us summarise these lessons.

1. First, some general lessons about particular judgments and theoretical principles. Particular judgments concern specific situations, and do not explain their verdicts. Theoretical principles, on the other hand, are general: they concern a wide range of situations. Moreover, they can both entail and explain particular judgments. Furthermore, a theoretical principle can have implications for cases beyond those the agent has explicitly considered.

2. Second, some observations on initial conditions. We will require that the agent is initially only aware of particular judgments.

3. Third, some lessons about the process of reaching reflective equilibrium. The agent must be able to formulate theoretical principles through awareness growth. She must also be able to formulate new particular judgments (that concern new situations) through awareness growth.

4. Fourth and finally, some observations on the state of reflective equilibrium. In the case of categorical belief, we require that once she has reached reflective equilibrium, the agent's belief set consists of some number of particular judgments together with a theoretical principle that entails and explains all of those particular judgments.

4.5 JUDGMENTS AND PRINCIPLES

I shall present two accounts of particular judgments and theoretical principles. The first account represents an ideal in the sense that it is suitable only to agents who are capable of making maximally fine-grained moral judgments. Although you and I are not such agents, it is nevertheless instructive to study the proposal, because it will allow us to formulate a very straightforward notion of reflective equilibrium that can be applied to all three models of belief. The second account is suitable to agents with finite powers of discrimination.

4.5.1 *The Fine-Grained Account*

Let's begin with particular judgments. A natural way of capturing the way in which they are particular is to say that they concern specific choice contexts. Your decision to order steak took place in a specific choice context, and it is this context alone that the judgment is concerned with. Let us call a claim about the moral status of an option which does not say why the option has this status an *evaluative* moral claim. And let us call a moral claim which does say why the option has this status an *explanatory* moral claim. With this terminology in place, we can now state our first account as follows: a particular judgment is a

context-specific evaluative claim.

Let us now see how this account deals with the various examples of particular judgments given above. Recall that a moral claim is a set of reasons structures, where a reasons structure is a pair $R = \langle N, \succeq \rangle$ consisting of a normative relevance function and a weighing relation. The normative relevance function tells us, for every context K in some background set $\mathcal{K}$ of contexts which properties are normatively relevant in that context, and the weighing relation ranks the available options by ranking their properties. Let K be the context in which you told a white lie (so that $[K]$ is the set of options available in K), and let w be the option of telling a white lie. The claim that telling a white lie was permissible will then be defined as follows:

$$PW = \{R \in \mathcal{R} : \forall x \in [K] : N_R(w, K) \succeq_R N_R(x, K)\},$$

where N_R and $\succeq_R$ are the normative relevance function and weighing relation of reasons structure R. That is, the claim that telling a white lie was permissible corresponds to the set of all and only those reasons structures according to which no option is ranked as strictly better than the option of telling a white lie.

Second, theoretical principles. A natural way of capturing the way in which they are general is to say that they correspond to (singleton sets of) individual reasons structures. After all, a theoretical principle is more or less the same thing as a moral theory, and the whole point of the reasons structure framework was to give us a way of formalising moral theories. This gives us:

FINE-GRAINED ACCOUNT Particular judgments are context-specific evaluative claims and theoretical principles are singleton sets of reasons structures.

On this account, the particular judgments are certainly specific, and by construction they do not cite reasons for their verdicts. It captures the particular character of the judgment by letting it concern just one choice context. And it captures the evaluative (as opposed to explanatory) nature of the judgment by including all reasons structures that entail the verdict in question. The various reasons structures in PW will identify different properties as normatively relevant. By not ruling out any of those reasons structures, we don't commit ourselves to any claims about which properties are normatively relevant and which ones are not.

It is straightforward to see how this account generalises to the other examples of particular judgments. The claim that it was impermissible to order steak for lunch will be the set of all and only those reasons structures according to which some other option is ranked strictly above the option of ordering steak in the relevant context. The claim that it would be obligatory to save the child drowning in the pond will be the set of reasons structures according to which this option is ranked strictly above every other option available in the context. And the claim that donating to the Against Malaria Foundation was better than donating to Oxfam will be the set of reasons structures that rank the former option strictly above the latter. And so forth.

The theoretical principles are certainly general, and clearly able to entail and explain particular judgments. They are general because they tell us, for any choice context, which properties are normatively relevant in that context, and they explain particular judgments by citing the normatively relevant properties in virtue of which those judgments hold. And they can also have implications for cases the agent has not explicitly considered. Furthermore, it is clear that an agent who begins with only particular judgments can formulate theoretical principles conceived of as single reasons structures via awareness growth.

The account also captures the state of reflective equilibrium nicely. In equilibrium, the agent's belief set will contain some number of particular judgments as well as a theoretical principle that entails and explains all of those judgments.

Problems for The Fine-Grained Account

However, the trouble with this account is that it makes both particular judgments and theoretical principles much too fine-grained.

Consider first particular judgments. According to the present proposal, particular judgments concern specific choice contexts. But choice contexts are extremely fine-grained. A context can be specified as a collection of sets of properties, with each set of properties representing an option available in the context. This means that whenever there is some discernible difference between two situations, they will correspond to different contexts. For example, if one situation takes place on a Tuesday and another on a Wednesday, they will correspond to different contexts, even if they are exactly alike in all other respects. In the former, all options (or to be more exact, all option-context pairs) will have the property of taking place on a Tuesday, whereas in the latter they will have the property of taking place on a Wednesday.

Given that contexts are so extremely fine-grained, how is it that an agent's particular judgment manages to latch on to one specific context rather than another? If the judgment in question concerns an actual situation the agent has encountered, we might be able to tell a causal story. When you judge that it was impermissible to order steak yesterday, this judgment was caused by the situation in question, and this situation corresponds to a single context. So here we might say that your judgment manages to latch on to a specific context in virtue of being causally related to an actual situation that instantiates that context. But whatever we may think of this response, it is unavailable if the particular judgment in question concerns a hypothetical situation rather than an actual one. If an agent forms a particular judgment about Singer's hypothetical scenario of the child drowning in a pond, there is no actual situation with which her judgment is causally connected. And moreover, the hypothetical scenario is underdescribed in various ways. This means that any individual context will contain a lot more detail than is present in the description of the case, and it would therefore be arbitrary to single out one context rather than

another.

Now, we could try to solve the problem as follows. Two contexts are distinct just in case there is some difference in terms of their properties. But we could reduce the number of different contexts by saying that the only difference-making properties are those that are possibly normatively relevant. For example, the day of the week is clearly not a normatively relevant property, so any two contexts that differ only in this respect will count as the same context. On this proposal then, two contexts are distinct just in case there is some difference in terms of their possibly normatively relevant properties.

Of course, this response lets the notion of a possibly normatively relevant property do a lot of work. The more properties we rule out as not possibly normatively relevant, the more plausible it becomes to say that particular judgments concern specific contexts. At the same time, the more properties we rule out, the more restrictions we impose on what a moral theory could look like. And given that our concern is ultimately to model agents who are morally uncertain, we should like to get by with as few such restrictions as possible.

Consider now theoretical principles on this account. Individual reasons structures are similarly much too fine-grained to play the role of theoretical principles. As you will recall, a reasons structure consists of a normative relevance function and a weighing relation. The normative relevance function tells us, for any possible context, what the normatively relevant properties are in that context, and the weighing relation tells us how to rank all bundles of properties. So in believing in a single reasons structure, an agent commits herself to definite beliefs about what would be normatively relevant in every possible context, and about how to rank options in every possible context.

According to this picture, what happens in the search for reflective equilibrium is that the agent begins with some particular judgments and then in formulating a theoretical principle she refines her algebra so that it now includes a maximally specific singleton set of a reasons structure. On what basis can we say that it is this reasons structure she believes in, rather than one of the countless other reasons structures that account equally well for her particular judgments and differ only on cases she is unaware of?

If the theoretical principle she believes in is of a very simple kind, such as total hedonistic utilitarianism, it is perhaps easier to answer this question convincingly. Here we can tell the following story: in reflecting on her particular judgments, the agent comes to realise that many or most of them track facts about total happiness: she tends to judge that one option is better than another if it leads greater happiness overall. The correlation need not be perfect, of course: in some cases she may intuitively think that an act is wrong even though it would lead to greater happiness overall. But through deliberation and philosophical reflection she gradually gives up these anti-utilitarian judgments and endorses total hedonistic utilitarianism. The fact that the theory is comparatively simple makes it less implausible to think that the agent believes

(in some sense of the word) its implications even for cases she has not considered at all.

Of course, even a comparatively simple theory such as this one has several free parameters. Do wanting and liking both count as pleasures? How do we measure pleasure? Which risk attitude is correct? And so forth. Presumably an agent can count as believing total hedonistic utilitarianism without having settled all such questions. But if so, we should not construe her as believing in a single reasons structure.

In summary, although this account is not suitable to agents like ourselves, it is nevertheless instructive to study it, because in a sense it represents an ideal. A logically omniscient agent with unlimited powers of discrimination would be able form particular judgments that concern specific choice context, and believe maximally specific theoretical principles construed as individual reasons structure. But for the rest of us, this is not always feasible.

4.5.2 *The Coarse-Grained Account*

Particular Judgments

According to our second account, particular judgments concern not specific contexts, but rather types of contexts, where a context type is given as a set of contexts. If an agent forms the judgment that some option is permissible in a given context type, she believes that it is permissible in all contexts of this type.

To illustrate, consider the standard trolley case. A runaway trolley is heading down a track towards five people. If you don't do anything, the trolley will hit and kill those five people. If you pull a lever, the trolley will be switched onto a different track, where it will hit and kill one person. Suppose an agent considers this case and forms the judgment that pulling the lever is obligatory. On the present proposal, this judgment concerns some set of contexts. In all of these contexts, the two available options are pulling the lever and not pulling the lever. How can we determine which contexts belong in this set and which ones don't? Put differently, how can we determine which contexts count as instances of the standard trolley case and which ones do not?

For some properties, it is determinate whether the options in the standard trolley case have them or not. For example, it is determinate that the option of pulling the lever has the property of leading to exactly one death. This means that any context in which the option of pulling the lever has the property of leading to any other number of deaths cannot be an instance of the standard trolley case. So any such context will not be an element of the set of contexts that represents the case.

For other properties, it is indeterminate whether the options in the standard trolley case have them or not. For example, it is indeterminate what temperature it is. So a context in which it is 20°C and a context in which it is 22°C may

both be instances of the case.

For yet other properties, it is determinate whether the options have them or not, even though they are not explicitly mentioned in the description of the case. For example, we can assume that pulling the lever does not, in addition to the one death, also lead to one person suffering a headache. And in general, whenever we describe a hypothetical scenario, it is assumed that all possibly relevant information is mentioned explicitly.

For now, I will simply assume that a particular judgment concerns some set of contexts, without specifying what determines which contexts belong in that set and which ones do not.

Recall the two main features of particular judgments that we identified at the outset: that they are specific rather than general, and that they are evaluative rather than explanatory. The second account captures both of these aspects. Granted, given that particular judgments now concern sets of contexts rather than individual contexts, they will be less specific than they were on the first account. But of course, the whole reason for going beyond the first account was that it made particular judgments too specific for realistic application. On the present proposal, those judgments are still specific in the sense that matters. An agent who judges that pulling the lever is obligatory in the standard trolley case does not thereby commit herself to any verdict about the bridge variant of the trolley case.

The present proposal also captures the way in which particular judgments are evaluative rather than explanatory. However, the distinction is somewhat less clear cut than it was for the first account, for the following reason. In a context type, as we have seen, some properties are determinately present, some are determinately absent, and for others it is indeterminate whether they are present or absent. So when an agent forms a particular judgment about a given context type, she implicitly commits herself to thinking that the indeterminate properties do not make a difference for the particular judgment. This is so because the agent forms the judgment in question regardless of whether these properties are present or not.

To illustrate, consider the following case. Let the context type be given as $\mathbf{K} = \{K_1, K_2\}$. There are two options, O_1 and O_2. In both K_1 and K_2, O_1 has property P_1 and O_2 has property P_2. In K_2, option O_2 additionally has property P_3. So in context type $\mathbf{K}$, it is determinate that O_1 has P_1 and that O_2 has P_2 but indeterminate whether O_2 has P_3. Suppose now that an agent forms the particular judgment that O_1 is better than O_2 in K. This doesn't mean that the agent implicitly judges that P_3 is normatively irrelevant. Consider a reasons structure according to which $N(K_1) = \{P_1, P_2\}$ and $N(K_2) = \{P_1, P_2, P_3\}$. If it has a weighing relation according to which $\{P_1\} > \{P_2\}$ and $\{P_1\} > \{P_2, P_3\}$, it will still accurately represent the agent's particular judgment. Hence what is ruled out is not that P_3 is normatively relevant, but only that its presence or

absence is not sufficient to change the way in which the two options are ranked.

Hence in making a particular judgment, the agent does not commit herself to any claim about which of the determinately present properties are normatively relevant. Nor does she commit herself to believing that the indeterminate properties are irrelevant, although she is forced to think that they are in a sense inert. But all of this means that a wide range of explanations, in terms of normatively relevant properties, can be provided for any given particular judgment. Therefore, particular judgments are still evaluative rather than explanatory.

On this account, the claim that telling a white lie was permissible will then be given as follows, where **K** is the set of contexts that belong to the context type in question:

$$PW = \{R \in \Omega : \forall K \in \mathbf{K} \forall y \in [K] : N_R(w, K) \succeq_R N_R(y, K)\}.$$

Theoretical Principles

According to the coarse-grained account of theoretical principles, they are sets of reasons structures that could, in reflective equilibrium, entail and explain all of the agent's particular judgments. Which set? Consider again the case of total hedonistic utilitarianism. As we saw, this theory is underdetermined in some ways. The fact that the phrase "total hedonistic utilitarianism" does not determine a unique reasons structure suggests a slightly different account from the first. We can think of the questions about whether both wanting and liking count as pleasure, and so forth as different ways of making the meaning of "total hedonistic utilitarianism" more precise. And then we can think of an agent who believes in total hedonistic utilitarianism without having settled on any particular way of making it maximally precise as believing in something more coarse-grained than a single reasons structure, namely the set of all admissible precisifications of total hedonistic utilitarianism.

This account captures the way in which theoretical principles are explanatory, although the notion of explanation is perhaps somewhat less straightforward than on the first account. A single reasons structure will tell us, for any given context, which properties are normatively relevant in that context. It is therefore able to explain particular judgments in terms of those normatively relevant properties. More specifically, if the agent judges that option O is obligatory in context type **K**, a reasons structure is able to explain this judgment by (i) telling us what the normatively relevant properties of the various options available in this context are, and (ii) telling us that the bundle of properties that corresponds to option O is ranked strictly higher than every other available option by the weighing relation.

By contrast, given that we are now construing theoretical principles as sets of reasons structures rather than individual ones, it may happen that those reasons structures offer different normatively relevant properties or different

weighing relations in explanation of the the same particular judgment. This means that the notion of explanation is perhaps less clear-cut, because it may be indeterminate in virtue of what set of normatively relevant properties a given particular judgment holds. Consider again the various precisifications of total hedonistic utilitarianism. "Pleasure" for these different theories will refer to slightly different things. We can think of these as various different pleasure properties: $pleasure_1$, $pleasure_2$, etc. So an action will be permissible either in virtue of leading to the most $pleasure_1$, and the most $pleasure_2$, etc. If two precisifications give different verdicts about some context, then insofar as she is aware of that context, she will not form a determinate judgment about this context herself, at least not if she is in reflective equilibrium. For if she did form a judgment, then she could rule out as inadmissible those precisifications that contradict this judgment.

The proposal also allows us to say that a theoretical principle can have implications for cases the agent has yet to consider. For example, suppose I take myself to be committed to utilitarianism and that I come across, for the first time, some purported counterexample to utilitarianism, say Foot's case of the surgeon who cuts up a healthy person in order to distribute their organs and save the lives of five other people. In considering this case, I come to see that my theoretical principle has implications I had not previously recognized, namely that it would be obligatory for the surgeon to cut up the healthy person. Indeed, I may even come to doubt my theoretical principle in the light of these implications.

Problem
The main problem with this account is that it's unclear how we determine which particular set of reasons structures correspond to a given theoretical principle. In other words, what determines the admissible precisifications of "total hedonistic utilitarianism"? The problem becomes especially pressing in light of the fact that a theoretical principle can have implications that the agent hasn't recognized. If the agent has not yet formed a judgment about some case because she is not aware of it, then how can we tell whether her principle implies one verdict rather than another?

A natural suggestion is to give a dispositional account of what it is for an agent to believe a theoretical principle. More specifically, we can say that an agent believes a theoretical principle in case she is disposed to form particular judgments in accordance with the principle. Given that she can be disposed to behave in some particular way even in cases she is not aware of, this allows us to explain which specific theoretical principle the agent believes.

However, this suggestion ignores the fact that agents may sometimes, as a result of becoming aware of a new case, come to *reject* the theoretical principle they previously accepted. Suppose an agent has formulated a theoretical principle and that she comes across a new case she had not previously considered. How can we distinguish between the following two cases?

1. She believes theoretical principle T_1 and forms the particular judgment J about the new case. T_1 entails J, so she can hold on to her original principle.

2. She believes theoretical principle T_2 and forms the particular judgment J about the new case. However, J is inconsistent with T_1, and she must therefore give up her original theoretical principle.

For example, how do we distinguish between the case in which the agent comes to reject utilitarianism in light of its implications in Foot's example, and the case in which she instead believed in some threshold deontological view according to which rights violations are only permissible when the consequentialist stakes are extremely high?

Given that she in both cases *in fact* forms the judgment that it would be impermissible for the surgeon to cut up the healthy person, it seems we should say that she has the disposition to form this judgment in both cases. But if so, then the dispositional solution is unable to account for the possibility that the agent may initially believe in utilitarianism and then come to reject it in light of its implications.

In response, we might suggest that one should consider a broader class of dispositions than just dispositions to form particular judgments. For example, if the agent came to reject utilitarianism, she will then be disposed to say things like "I used to be a utilitarian but now I see the error of my ways." Or we might suggest that an agent believes a theoretical principle just in case she is under *normal* circumstances disposed to accept its verdicts, and then specify a suitable notion of normality which rules out the relevant cases. However, I will consider these suggestions in any more detail.

On a general level, the problem appears to be an instance of a familiar skeptical worry about rule-following and meaning. Kripke (1982) takes Wittgenstein (1953/2009) to present a certain sceptical challenge. Consider the following question: in virtue of what is it that I mean *addition* when I use the '+' symbol, and not some other arithmetical operation? It is true that all of my past usage of the symbol has been consistent with addition. But suppose that I've never calculated "$68 + 57$" before. I perform the calculation and arrive at "125." How can I be sure that that my answer is correct?

It is true that adding 68 to 57 gives 125. But how can I be sure that when I used the '+' symbol in the past, I intended it to mean not addition but *quaddition*, which we can denote $\oplus$. Quaddition is defined as follows: $x \oplus y = x + y$ if $x, y < 57$; otherwise $x \oplus y = 5$. Nothing about my past usage rules out the possibility that I intended for '+' to mean quaddition all along. So what else can account for the fact that I mean addition and not quaddition? As we have seen, dispositional accounts face difficulties. But if there is nothing in virtue of I meant addition rather than quaddition, it seems there can be no fact of the matter as to which of the two I meant.

In our case then, the rule is not addition, but rather some theoretical principle T that the agent takes herself to be committed to. She has applied this principle to various cases in the past, but now she has become aware of a new case. Perhaps all along she has not meant to refer to T but rather to some other theoretical principle T' which agrees with T on all past cases, but disagrees on the new one. So how can she be sure whether the judgment she forms about the new case is consistent with her theoretical principle?

This is not the place for a detailed examination of possible solutions to this skeptical worry. I will assume that some solution exists, and formulate an account of reflective equilibrium that is suitable to theoretical principles construed as sets of reasons structures. Insofar as we take this to be a problem, the problem is not particular to our choice of formal framework. Any account of reflective equilibrium which allows that agents can believe principles that have implications beyond those they have recognized will face the same problem.

4.6 BAYESIAN REFLECTIVE EQUILIBRIUM

I have presented an informal survey of reflective equilibrium, and now given definitions of particular judgments and theoretical principles. It's time to put all the pieces together and provide a formal account of reflective equilibrium. I shall mainly be doing so in terms of the fine-grained account of judgments and principles, but as we shall see, my notion of reflective equilibrium can easily be extended to the coarse-grained account. Given that the account of reflective equilibrium that I propose can be adapted to each of the three models of beliefs we have considered, let us consider these in turn.

4.6.1 *Categorical Reflective Equilibrium*

In the case of categorical belief, an agent's belief set $\mathcal{B}$ is in reflective equilibrium just in case it consists of some theoretical principle T together with some set $\mathcal{J}$ of particular judgments such that $T \subset J$ for each $J \in \mathcal{J}$. In other words, the theoretical principle *entails* all of the particular judgments. Moreover, T explains these particular judgments by citing the normatively relevant properties in virtue of which this judgment holds. Recall that a particular judgment is a context-specific evaluative claim, and that a theoretical principle is a single reasons structure. So for example, if J is the judgment that option O_1 is better than option O_2 in context K, then T explains J by (i) telling us what the normatively relevant properties of the options are in the given context (i.e. $N(O_1, K)$ and $N(O_2, K)$ respectively) and (ii) telling us via the weighing relation $\succeq$ that (in virtue of their respective properties) O_1 is better than O_2.

Therefore, an agent is in categorical reflective equilibrium just in case (i) her belief set contains some theoretical principle that entails all her moral beliefs, and (ii) her belief set is coherently extendable. As should be clear, this definition applies equally well if we use the coarse-grained for judgments and principles instead. The particular judgments will now concern context types rather than individual contexts, and the theoretical principle will now be a set of reasons

structures rather than an individual one. But the theoretical principle must still entail all her moral beliefs, and her belief set must still be coherently extendable.

Once we have this simple model in place, we can relax some of the assumptions. For example, we can consider what happens when the agent's σ-algebra consists not just of judgments and principles and combinations thereof, but rather of any kind of moral claim. We are thus considering a σ-algebra that contains arbitrary sets of reasons structures. However, the same notion of reflective equilibrium is still applicable: we say that an agent is in reflective equilibrium just in case her belief set contains some theoretical principle that entails every element of her belief set.

We have not imposed any constraints on which theoretical principles may count as explanatory. As long as it entails all of the judgments she holds in reflective equilibrium, an agent may take any reasons structure to be a satisfactory explanation of those judgments. But this means that agents may regard reasons structures that strike us as utterly bizarre and unsystematic as providing a satisfactory account of their moral judgments. For example, according to some reasons structures, murder is impermissible, except if the day of the month is evenly divisible by 5. Clearly such a principle cannot be an adequate account of an agent's moral judgments. However, if they strike us as bizarre, they will presumably strike the agent as bizarre as well, and she will not consider them to provide a plausible explanation of her judgments. There is certainly much to be said here, but we should not expect a formal account of reflective equilibrium to tell us which features of theoretical principles make them eligible to play an explanatory role. That is a task for substantive moral inquiry.

Indeed, moral particularists such as Dancy (2004) maintain that although there are various true moral claims, there are no true moral principles, at least if moral principles are required to exhibit the sort of regularity that the bizarre principle we just considered did not have. However, as long as she tells us both her judgments about various cases and the reasons for those judgments, we can still construe the particularist as believing in a theoretical principle in the technical sense that there is some reasons structure that entails all of her judgments. But of course, such a reasons structure would lack various paradigmatic explanatory virtues. On the other hand, a particularist may not have much interest in the present project to begin with:

> for reflective equilibrium (whether wide or narrow) one is required to establish a stable balance between particular judgments and general principles, and this is something that no particularist is going to see any need for (Dancy 2004:153).

4.6.2 *Relational Reflective Equilibrium*

Recall that the basic judgments of relational belief are comparative confidence judgments of the form "A is at least as likely as B," which we write as $A \trianglerighteq B$. A relational doxastic state can be modelled as a relational belief set $\mathcal{B}_{\trianglerighteq} \subseteq \mathcal{F} \times \mathcal{F}$ containg all the comparative confidence judgments she accepts. To make it specific, suppose that we accept de Finetti's axioms on qualitative probability from section 3.4.2. What does reflective equilibrium look like for an agent who makes qualitative probability judgments? Her belief set $\mathcal{B}_{\trianglerighteq}$ consists of comparative claims, where the claims being compared are moral claims. More specifically, they are either theoretical principles or particular judgments. The first thing to note is that it's no longer possible to require that agents in reflective equilibrium must believe a theoretical principle that entails all of their other moral judgments by logical consequence. By construction, there is no theoretical principle T that she believes categorically, but rather a set of theoretical principles $\mathcal{T} = \{T_1, T_2, \dots\}$ ordered by the qualitative probability relation.

However, given that we are not concerned with categorical belief, logical consequence is no longer the appropriate consequence relation. Instead, we should appeal to consequence with respect to the qualitative probability axioms. This way, we can say that an agent is in relational reflective equilibrium just in case her relational belief set is (i) coherently extendable with respect to the qualitative probability axioms, and (ii) contains claims that concern both particular judgments and theoretical principles. The second condition is there to rule out agents who have not yet attempted to systematize their particular judgments by formulating theoretical principles. Each theoretical principle is still such that, were she to believe it categorically, it would be able to entail some set of categorically believed particular judgments. But as we are dealing with relational belief, we have to settle for qualitative probability consistency instead.

4.6.3 *Bayesian Reflective Equilibrium*

Finally, the basic judgments of quantitative belief are of the form "I believe A to degree x." Among quantitative models of belief, we are of course mainly interested in the probabilistic one. We model an agent's probabilistic doxastic state as a probabilistic belief set $\mathcal{B}_P \subseteq \mathcal{F} \times \mathbb{R}$.

The general strategy for how to define a notion of reflective equilibrium suitable to the corresponding category of belief should now be familiar. An agent is in probabilistic reflective equilibrium just in case her probabilistic belief set is (i) coherently extendable with respect to the probability axioms, and (ii) contains claims that concern both particular judgments and theoretical principles.

Suppose $\mathcal{F}$ initially consists only of some set of particular judgments $\mathcal{J} = \{J_1, J_2, \dots\}$ (closed under complements and countable unions). These particular judgments are context-specific evaluate claims. In formulating a theoretical principle, her algebra grows from $\mathcal{F}$ to $\mathcal{F}^+$, where $\mathcal{F}^+$ now contains some theoretical principle T. Is this awareness growth by refinement or awareness

growth by expansion? Given that we want T to entail each of the particular judgments, it seems that it must be refinement. For in order for T to entail J, we must have that $T \subset J$. Hence given that $J \subset \Omega$, we must also have that $T \subset \Omega$. Therefore, T cannot have been formulated through awareness growth.

Recall now the principle of Awareness Rigidity, which says that for any $A \in \mathcal{F} \cap \mathcal{F}^+, P^+(A \mid \Omega) = P(A)$. In cases of refinement, of course, $P^+(A \mid \Omega) = P^+(A)$. So Awareness Rigidity entails that when an agent formulates theoretical principles, her credences in the particular judgments must remain unchanged. But that seems to run counter to the whole idea of the reflective equilibrium process, namely that reflecting on theoretical principles may cause the agent to abandon or revise some of her earlier judgments. So it seems Awareness Rigidity is not a plausible constraint on reflective equilibrium.

However, we could instead model this part of the reflective equilibrium procedure as a two-step process. First, the agent's awareness grows from $\mathcal{F}$ to $\mathcal{F}^+$, and she initially assigns credences in accordance with Awareness Rigidity. At this point, she is simply trying formulate theoretical principles that might be able to account for her particular judgments. Therefore, her credences should not change at this stage, for it is those very credences that the theoretical principles are supposed to account for. However, at a later stage the agent may reflect on the theoretical principles and as a result revise her credences in the particular judgments, perhaps in accordance with Jeffrey conditionalization. Of course, this still leaves the process rather unconstrained. But perhaps that is as it should.

How does this proposal relate to the distinction between narrow and wide reflective equilibrium? Nothing in our framework makes it strongly biased towards the initial judgments: an agent could in principle end up in a reflective equilibrium in which she rejects all of her initial judgments. So our notion of reflective equilibrium is not narrow in this sense. On the other hand, we have not required that she first carefully consider a set of prominent alternative moral principles, nor have we required that she study how her favoured principle squares with well-established facts of human psychology, etc., so our notion is not a wide one either. But it can serve as the starting point for an account of wide reflective equilibrium. For example, we might require that for wide reflective equilibrium the agent must have considered some set $\mathcal{T}$ of prominent alternative theoretical principles. But I will leave such extensions for another day.

You might wonder if this proposal is sufficiently stable and satisfying to really count as reflective equilibrium. For example, an agent may be in reflective equilibrium even though her credence is evenly divided among ten theoretical principles, each one very different from the next. Should we wish to, we can make various amendments to this basic account of Bayesian reflective equilibrium. For example, we might for example take inspiration from Leitgeb's stability theory of belief and require that in order for an agent to be in reflective equilibrium, there must in addition be some theoretical principle in which she

has stably high credence, where a credence is stably high if it would remain above some threshold were the agent to conditionalize on any one of a suitably defined set of propositions. But I shall again leave extensions for the future.

4.7 CONCLUSION

This chapter has covered a lot of ground, so a recap is in order. Our starting point was probabilism with respect to moral uncertainty, and our question was how probabilistic credences in moral claims should be updated over time. We initially considered some cases in which (Jeffrey) conditionalization did appear to be an adequate update rule for degrees of belief in moral claims, such as relying on moral testimony, treating one's own emotions as evidence, or consulting the historical track record of claims made on behalf of some particular moral theory.

However, I suggested that not all paradigmatic cases of moral reasoning can be modelled as instances of (Jeffrey) conditionalization. Our first example concerned thought experiments. I argued that many thought experiments play the role of making agents realize that a moral principle they previously accepted has implications they had not initially appreciated. Others play the role of forcing agents to make a judgment about a case they had not previously considered. I argued that the best way to capture both of these aspects is to allow agents to undergo awareness growth.

I then suggested that awareness growth is also necessary if we are to give a Bayesian account of reflective equilibrium. I examined two accounts of particular judgments and theoretical principles. According to the fine-grained account, particular judgments were context-specific evaluative claims, and theoretical principles were individual reasons structures. Although this account makes both judgments and principles implausibly fine-grained, we nevertheless found it instructive as an objecte of study, because it affords us the useful idealization of an agent who is able to make maximally fine-grained moral judgments.

According to the coarse-grained account, particular judgments were evaluative claims concerning context types, where a context type is a set of contexts. A theoretical principle was a set of reasons structures whose extension is determined by the agent's disposition to accept or reject various particular judgments and other moral claims.

Using these accounts of particular judgments and theoretical principles, we proposed a notion of reflective equilibrium suitable to each of the three types of belief models we have considered. The general account goes as follows: an agent is in reflective equilibrium just in case she has formulated an adequate (set of) theoretical principles to account for her particular judgments, and her belief set is coherently extendable. Plugging in the different notions of coherent extendability, this gives us a definition of reflective equilibrium for each of

categorical, relational, and probabilistic belief.

We have seen that the constraints on rational belief change are relatively weak in the case of awareness growth. Two agents may start with the same probability function P and undergo the same episode of awareness growth from $\mathcal{F}$ to $\mathcal{F}^+$ and yet end up with radically different probability functions P_1^+ and P_2^+ afterwards. If I am right that awareness growth plays a more prominent role in moral epistemology, it would seem to follow that our moral credences are less constrained over time.

Part III

CODA

5

CONCLUSION

5.1 SUMMARY

Our starting point was the traditional Bayesian picture according to which the structure of credence is given by probabilism, its objects are descriptive propositions about the empirical world, and its dynamics are those of (Jeffrey) conditionalization. We have explored three variations on this Bayesian theme. Although the three essays concerned somewhat different topics, some general lessons nevertheless emerge. But let us first briefly rehearse what I take to be the main takeaways from each essay individually.

5.1.1 *Imprecise Bayesianism and Global Belief Inertia*

In chapter 2, we explored a variation in the structure of credence. Specifically, we considered imprecise Bayesianism, which models the agent's credal state as a set of probability functions rather than a single one. We saw that imprecise Bayesianism is often motivated by the evidentialist thought that our credences shouldn't be more precise than is called for by the evidence. We considered Joyce's (2005, 2010) way of spelling out this evidentialist thought in more detail, and I argued that we should reject his view because it implausibly entails:

GLOBAL BELIEF INERTIA For any proposition A, a rational agent will have a maximally imprecise credence in A unless her evidence logically entails either A or its negation.

However, the conclusion to draw is not that we should reject imprecise Bayesianism in its entirety. Nor is it even that we should reject all forms of evidentially-motivated imprecise Bayesianism. The specific aspect of Joyce's imprecise Bayesianism which the objection concerns is the fact that it makes imprecision of a particular kind rationally required in certain evidential situations. For example, in UNKNOWN BIAS Joyce required that if you have no evidence at all as to the bias of a coin, then your credence that it will come up heads should be maximally imprecise.

By contrast, on a view according to which imprecise credences are rationally permissible rather than rationally required, global belief inertia does not follow. Moreover, such a permissive imprecise Bayesianism could also be given an evidentialist motivation. For example, one possible view is that if $\mathcal{P}$ is the set of all probability functions that are compatible with some given body of

evidence $\mathcal{E}$, then any subset of $\mathcal{P}$ is a rationally permissible credal state to adopt in light of $\mathcal{E}$. Granted, this view still makes global belief inertia rationally permissible. However, I think that making global belief inertia rationally permissible is significantly less of a problem than is making it rationally required. Moreover, ruling out that global belief inertia is rationally permissible appears to be a much more difficult task than ruling out that it is rationally required. An agent who suffers from global belief inertia is more or less an inductive skeptic: she is essentially unwilling to draw any conclusions from her observations. Therefore, in order to show that global belief inertia is irrational, we must show that inductive skepticism is irrational.

But even ruling out this strict form of inductive skepticism would not be enough. We also want credences to change quickly enough in response to incoming evidence. In practice, an agent who after having seen a million heads and zero tails has only become moderately more confident that the coin is either two-headed or otherwise biased towards heads is barely any better than an agent whose confidence does not change at all. We want agents to be able make reasonable inductive inferences based on the limited amount of evidence they will receive in their finite lifespan. Hence we would have to show not just that credences are responsive to the evidence, but that they are sufficiently responsive to that evidence. But of course now the question becomes one of how the rate at which one learns and commits oneself to definite beliefs should be traded off against the risk of thereby committing oneself to false beliefs. And here it seems that reasonable people may disagree. Indeed, we can view these as instances of the two Jamesian goals of belief, pulling in different directions:

> There are two ways of looking at our duty in the matter of opinion,–ways entirely different, and yet ways about whose difference the theory of knowledge seems hitherto to have shown very little concern. We must know the truth; and we must avoid error,–these are our first and great commandments as would be knowers; but they are not two ways of stating an identical commandment [...] 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. We may regard the chase for truth as paramount, and the avoidance of error as secondary; or we may, on the other hand, treat the avoidance of error as more imperative, and let truth take its chance (James 1896, Section VII).

If my belief that the coin will land heads is maximally imprecise, $P(H) = [0, 1]$, there is no risk of error, because I have not committed myself to any definite belief about the outcome of the coin flip. The more precise I make my belief, the greater risk I run of error. At the same time, of course, the greater chance I have of committing myself to an accurate more precise belief. Therefore, insofar as people may rationally disagree over the relative importance of the two goals of belief, it seems that they may rationally differ in the precision of the credences that they adopt in response to a given body of evidence.

5.1.2 *Moral Uncertainty and Arguments for Probabilism*

In chapter 3, we explored a variation in the object of credence. We considered whether the three standard arguments for probabilism—representation theorem arguments, Dutch book arguments, and accuracy arguments—can also establish this conclusion when the objects of credence are moral claims rather than empirical claims. I first proposed that we can model moral claims as sets of reasons structures, in analogy with the standard possible worlds semantics for descriptive propositions. We then examined how the various arguments for probabilism fare when the σ-algebra consists of sets of reasons structures. I argued that decision-theoretic representation theorem arguments are more problematic in the case of moral uncertainty than in the case of empirical uncertainty, because in the former case they commit us to thinking that all moral theories are intertheoretically comparable. Epistemic representation theorem arguments do better, but I suggested that rationally permissible violations of completeness may be more widespread in the case of moral uncertainty than in the case of descriptive uncertainty. I then argued that if you take the depragmatized Dutch book argument to make a plausible case for probabilism with respect to descriptive uncertainty, you should also find it persuasive with respect to moral uncertainty, at least provided that it is not irrational to evaluate bets in a way that is contrary to one's moral convictions. Finally, I suggested that moral uncertainty does not pose any particular problems for accuracy arguments.

Overall, we found the case for probabilism to be somewhat less clear-cut for moral uncertainty, but on balance still fairly good, especially if we allow for imprecise credences. We can now see why the argument against imprecise credences in the previous essay and the argument in favour of them in this one are not in tension with one another: the former is an argument against a view that makes a certain kind of imprecision rationally required, whereas the latter is an argument in favour of a view that makes it rationally permissible. Finally, we did not engage at sufficient length with non-cognitivism. The arguments we considered make most intuitive sense on a cognitivist understanding, and any conclusions established should therefore be read as holding conditional on cognitivism. It may be that there are sound non-cognitivist readings of all of these arguments, but we have not attempted to provide them.

5.1.3 *Bayesian Moral Epistemology and Reflective Equilibrium*

In chapter 4, we explored a variation in the dynamics of credence. Specifically, we considered the role of awareness growth in giving a Bayesian account of reflective equilibrium. Our starting point was the assumption that something like probabilism holds for moral uncertainty. I argued that there are some cases in which (Jeffrey) conditionalization does appear to be the correct update rule for our moral credences, but that many paradigmatic cases of moral learning are not covered by this rule. In particular, I argued that if we wish to make room for reflective equilibrium in our Bayesian moral epistemology, then we

must allow for awareness growth.

I then proposed two accounts of particular judgments and theoretical principles. According to the fine-grained account, a particular judgment is a context-specific evaluative claim and a theoretical principle is a single reasons structure. According to the coarse-grained account, particular judgments instead concern sets of contexts, and theoretical principles are non-singleton sets of reasons structures. The first account is appropriate for idealized agents who are capable of making maximally fine-grained judgments, whereas the second account is appropriate for finite agents like ourselves.

The coarse-grained account ran into a difficulty concerning how to specify which particular theoretical principle an agent believes, given that such a principle is supposed to be able to have implications for cases beyond those the agent has explicitly considered. I noted that the general difficulty appears to be an instance of (Kripke's) Wittgenstein's skeptical worries about rule-following, and argued that far from being unique to the particular formal framework I have chosen to work with, any account of reflective equilibrium which allows that agent can believe a theoretical principle without having recognized all of its implications will be subject to the same difficulty. Given that this appears to be a central feature of the methodology, any plausible account of reflective equilibrium will therefore face the same problem.

I then gave a general account of reflective equilibrium: an agent is in reflective equilibrium just in case (i) she has formulated adequate theoretical principles, and (ii) her belief set is coherently extendable. An advantage of this account is that it applies equally well to all of the three models of belief we have considered. And it applies equally well on both the fine-grained and the coarse-grained account.

Finally, we saw that the constraints on rational credence appear to be much weaker in the case of awareness growth than in the case of ordinary updating. Therefore, insofar as awareness growth is indeed more central in moral epistemology than it is in descriptive epistemology, it follows that the rationality constraints on moral credences are correspondingly weaker.

5.2 LESSONS AND FUTURE DIRECTIONS

5.2.1 *Coherent Extendability*

In section 1.2, I introduced three models of belief: categorical, relational, and quantitative. Although we have not had much to say about categorical belief, we have extensively discussed both relational and quantitative belief. Epistemic representation theorems show us how relational belief and quantitative belief relate to one another. In particular, in section 3.4.2 we saw both how comparative confidence judgments relate to precise probability, and how they relate to imprecise probability.

If we don't take completeness to be a rationality requirement on comparative confidence judgments, then we will not take precise credences to be rationally required either. But we can still require that the comparative confidence judgments can be given an imprecise probability representation. If so, then the notion of coherent extendability can provide a natural account of the way in which imprecise Bayesianism is still distinctively Bayesian. In our discussion of epistemic representation theorems we argued that even if completeness is not rationally required of our comparative confidence judgments, it should still be rationally permissible. But if an agent does become fully opinionated, then the rationality assumptions guarantee that her beliefs can now be given a probabilistic representation.

However, some think that complete comparative confidence judgments are not always rationally permissible. I suggested that those who think so also tend to think that at other times complete comparative confidence judgments are in fact rationally permissible, and perhaps even rationally required. In particular, we might think that if all of an agent's evidence consists of chance propositions, and this evidence is extensive enough to determine, via the correct chance-credence principle, a unique value for each proposition in her σ-algebra, then she is rationally permitted to adopt a precise credence function. Furthermore, I suggested that any rational agent should at least in principle be able to learn some body of evidence that would rationalize precise credences. This means that we can replace coherent extendability with evidential extendability: an agent is evidentially extendable just in case she could learn some body of evidence that would rationalize precise credences without thereby becoming strictly inconsistent.

We also appealed to coherent extendability in our formal account of reflective equilibrium. In particular, I suggested that an agent is in reflective equilibrium just in case she has formulated suitable theoretical principles and her belief set is coherently extendable. Given that coherent extendability can be specified in different ways by specifying a different consequence relation, our concept of reflective equilibrium can be usefully applied to all three models of belief. Here too the notion allowed us to provide a natural account of what Bayesianism required of agents with incomplete attitudes.

5.2.2 *Awareness Growth*

Awareness growth played a central role in my account of moral epistemology. I introduced it by arguing that we need it to capture the role played by thought experiments, and then suggested that we also need it to capture the process of reflective equilibrium. Although I have emphasized the role of awareness growth in moral epistemology, I do not mean to imply that it is unimportant in the rest of epistemology. Indeed, I take it that any realistic application of Bayesian epistemology must allow for awareness growth, regardless of the subject matter. However, awareness growth is still a relatively unexplored phenomenon, and much remains to be done.

5.2.3 *Bayesian Moral Epistemology*

In considering the prospects for a Bayesian moral epistemology, we have been exploring largely uncharted territory. Granted, the notion of probabilistic credences in moral claims figures in many existing discussions of moral uncertainty. However, those discussions tend to focus on decision-making, and therefore neglect to consider the epistemology of moral uncertainty in any detail. Furthermore, they rarely if ever address the question of how our credences in moral claims should change over time as we acquire evidence and engage in moral reasoning, thereby leaving out a crucial component of the epistemology.

Why is it that we have a well-developed Bayesian scientific epistemology (i.e. Bayesian confirmation theory) but not a well-developed Bayesian moral epistemology? Perhaps part of the reason is that the precision of traditional Bayesianism appears almost comically ill-suited to the moral domain. It's absurd to imagine that an agent might have a credence of 0.345716 in utilitarianism, a credence of 0.257321 in Kantianism, and so forth. Although the same objection has been raised for Bayesian confirmation theory, here it appears to be if anything more forceful. But if I am right, then an imprecise moral epistemology allows us to reap the rewards of Bayesianism without paying the price of implausible precision. Moreover, this imprecise moral epistemology is still essentially a Bayesian moral epistemology, because given the requirement of coherent extendability, were the agent to become fully opinionated, she would thereby commit herself to precise credences. So although our moral epistemology is Bayesian, it still allows for the possibility that our moral beliefs may in general be less precise than our descriptive beliefs.

It emerged from our discussion of reflective equilibrium that a Bayesian moral epistemology may turn out to be less determinate than its descriptive counterpart along another dimension as well. In particular, I suggested that awareness growth may play a more central role in our moral reasoning than in our descriptive reasoning. If this is right, it follows that the dynamics of rational credence are less constrained when those credences concern moral matters, because there are fewer rationality constraints on updating by awareness growth than there are on standard forms of updating.

If our moral beliefs are both less determinate and less constrained by the evidence than are our descriptive beliefs, rational disagreement will be more widespread in the moral domain than in the descriptive. More generally, we might take it to indicate that developing true moral beliefs may be a rather difficult undertaking. In light of this, perhaps the appropriate attitude to take is one of moral humility.

For example, consider the argument from disagreement, which takes the fact that moral disagreement is so widespread to indicate that there is no moral fact of the matter. If I am correct that credences in moral claims are less constrained in how they change over time, this might suggest a new way of responding to

the argument, namely that the disagreement we observe may just as well be the result of different agents drawing different conclusions from the same episodes of awareness growth.

Of course, these are just the tentative conclusions of an initial exploration of Bayesian moral epistemology. It may be that further examination reveals it to be much more precise and constrained than I have suggested. And it may be that we wish to impose more substantive constraints on reflective equilibrium than those we have considered in our formal account. Indeed, the sketch of Bayesian moral epistemology that I have provided leaves many questions unaddressed. For example, how does the debate between subjective and objective Bayesianism play out in the case of moral uncertainty? Do we have reason to accept any of the further constraints on prior probability that were discussed in section 1.4? Are there any further constraints on priors that are particular to the moral domain? Once we begin to give Bayesian moral epistemology serious consideration, several interesting new avenues of investigation open up.

6

BIBLIOGRAPHY

Arnauld, Antoine and Nicole, Pierre (1662/1996). *Logic or the Art of Thinking*. Cambridge: Cambridge University Press. Translated by Jill Vance Buroker.

Ashbery, John (2002). Soonest Mended. In *Selected Poems*. Manchester: Carcanet Press, pp. 87–89.

Bertrand, Joseph (1889). *Calcul des probabilités*. Paris: Gauthier-Villars.

Blackwell, David and Dubins, Lester (1962). Merging of opinions with increasing information. *The Annals of Mathematical Statistics* 33:882–886.

Bolker, Ethan D. (1966). Functions resembling quotients of measures. *Transactions of the American Mathematical Society* 124(2):292–312.

BonJour, Laurence (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Bradley, Richard (2017). *Decision Theory with a Human Face*. Cambridge: Cambridge University Press.

Bradley, Richard and Stefánsson, H. Orri (2016). Desire, Expectation, and Invariance. *Mind* 125(499):691–725.

Bradley, Seamus (2015). Imprecise Probabilities. In Edward N. Zalta (*ed*.) *The Stanford Encyclopedia of Philosophy*.

Bradley, Seamus and Steele, Katie (2014). Uncertainty, Learning, and the "Problem" of Dilation. *Erkenntnis* 79(6):1287–1303.

Brier, Glenn W. (1950). Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review* 78(1):1–3.

Briggs, Rachael (2015). Foundations of Probability. *Journal of Philosophical Logic* 44(6):625–640.

Briggs, Rachael and Pettigrew, Richard (forthcoming). An Accuracy-Dominance Argument for Conditionalization. *Noûs*.

Bykvist, Krister (2017). Moral Uncertainty. *Philosophy Compass* 12(3).

Bykvist, Krister and Olson, Jonas (2009). Expressivism and Moral Certitude. *Philosophical Quarterly* 59(235):202–215.

—— (2012). Against the Being For Account of Normative Certitude. *Journal of Ethics and Social Philosophy* 6(2):1–8.

Cath, Yuri (2016). Reflective Equilibrium. In Herman Cappelen, Tamar Gendler and John Hawthorne (eds.), *Oxford Handbook of Philosophical Methodology*. New York: Oxford University Press.

Chalmers, David J. (2011). Frege's Puzzle and the Objects of Credence. *Mind* 120(479):587–635.

Christensen, David (2004). *Putting Logic in its Place: Formal Constraints on Rational Belief*. Oxford: Oxford University Press.

Conee, Earl and Feldman, Richard (2004). *Evidentialism: Essays in Epistemology.* Oxford: Oxford University Press.

Dancy, Jonathan (2004). *Ethics without Principles.* Oxford: Oxford University Press.

Daniels, Norman (2018). Reflective Equilibrium. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition).

de Finetti, Bruno (1931). Sul significato soggettivo della probabilità. *Fundamenta Mathematicae* 17:298–329.

——— (1951). La "logica del plausibile" secondo la concezione di Polya. In *Atti della XLII Riunioni.* Rome: Societa Italiana per il Progresso delle Scienze.

——— (1964). Foresight: Its Logical Laws, Its Subjective Sources. In Henry E. Kyburg and Howard E. K. Smokler (eds.), *Studies in Subjective Probability.* Huntington, NY: Robert E. Kreiger Publishing Co.

Diaconis, Persi and Zabell, Sandy L. (1982). Updating Subjective Probability. *Journal of the American Statistical Association* 77(380):822-830.

Dietrich, Franz and Christian List (2017). What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories. *The Philosophical Review* 126(4):421–479.

Drèze, Jacques H. and Rustichini, Aldo (2004). State-Dependent Utility and Decision Theory. In Salvador Barberà, Peter J. Hammond and Christian Seidl (eds.), *Handbook of Utility Theory*, Volume 2: Extensions. New York: Springer, pp. 839-892.

Earman, John (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory.* Cambridge, MA: The MIT Press.

Easwaran, Kenny (2013). Why Countable Additivity? *Thought: A Journal of Philosophy* 2(1):53-61.

Elga, Adam (2010). Subjective Probabilities Should Be Sharp. *Philosopher's Imprint* 10(5):1–11.

Enoch, David (2014). A Defense of Moral Deference. *The Journal of Philosophy* 111(5):229–258.

Eriksson, Lina and Alan Hájek (2007). What Are Degrees of Belief? *Studia Logica* 86(2):185–215.

Fine, Terrence L. (1973). *Theories of Probability*. New York: Academic.

Fishburn, Peter C. (1986). The Axioms of Subjective Probability. *Statistical Science* 1(3):335–358.

Gaifman, Haim (2004). Reasoning with Limited Resources and Assigning Probabilities to Arithmetical Statements. *Synthese* 140(1-2):97–119.

Gaifman, Haim and Snir, Marc (1982). Probabilities over rich languages, testing, and randomness. *Journal of Symbolic Logic* 47:495–548.

Gärdenfors, Peter and Sahlin, Nils-Eric (1982). Unreliable Probabilities, Risk Taking, and Decision Making. *Synthese* 53(3):361–386.

Gilboa, Itzhak and Schmeidler, David (1993). Updating Ambiguous Beliefs. *Journal of Economic Theory* 59:33–49.

Glymour, Clark (1980). *Theory and Evidence.* Princeton: Princeton University Press.

Good, Irving John (1971). 46656 Varieties of Bayesians. *The American Statistician* 25(5):56–63.

Goodman, Nelson (1983). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Gracely, Edward J. (1996). On the Noncomparability of Judgments Made by Different Ethical Theories. *Metaphilosophy* 27(3):327–332.

Greaves, Hilary and Wallace, David (2005). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind* 115(459):607–697.

Guerrero, Alexander A. (2007). Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution. *Philosophical Studies* 136(1):59–97.

Gustafsson, Johan E. and Olle Torpman (2014). In Defence of My Favourite Theory. *Pacific Philosophical Quarterly* 95(2):159–174.

Hacking, Ian (1967). Slightly More Realistic Personal Probability. *Philosophy of Science* 34(4):311–325.

Hájek, Alan (2003). What Conditional Probability Could Not Be. *Synthese* 137(3):273–323.

——— (2008). Dutch Book Arguments. In Paul Anand, Prasanta Pattanaik and Clemens Puppe (eds.), *The Oxford Handbook of Rationality and Social Choice.* Oxford: Oxford University Press.

——— (2009). Arguments For–Or Against–Probabilism? In Franz Huber and Christoph Schmidt-Petri (eds.) *Degrees of Belief*. Springer.

——— (2012). Is Strict Coherence Coherent? *dialectica* 66(3):411–424.

——— (manuscript). Staying Regular? Available at `http://hplms.berkeley.edu/HajekStayingRegular.pdf`.

Harman, Elizabeth (2015). The Irrelevance of Moral Uncertainty. In Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics* 10.

Hausman, Daniel M. (2011). *Preference, Value, Choice, and Welfare*. Cambridge: Cambridge University Press.

Hawthorne, James (2008). Inductive Logic. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.

Hill, Brian (2013) Confidence and decision. *Games and Economic Behavior* 82:675–692.

Hills, Alison (2013). Moral Testimony. *Philosophy Compass* 8(6):552–559.

Hintikka, Jaakko (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic* 4(4):475–484.

Hudson, James L. (1989). Subjectivization in Ethics. *American Philosophical Quarterly* 26(3):221–229.

James, William (1896). The Will to Believe. *The New World* 5:327–347.

Jaynes, Edwin T. (1957). Information Theory and Statistical Mechanics. *The Physical Review* 106(4):620–630

——— (1973). The Well Posed Problem. *Foundations of Physics* 4(3):477–492.

——— (2003). *Probability Theory: The Logic of Science.* Cambridge: Cambridge University Press.

Jeffrey, Richard C. (1965). *The Logic of Decision*. Chicago: The University of Chicago Press.

Joyce, James M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science* 65(4):575–603.

——— (1999). *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

——— (2005). How Probabilities Reflect Evidence. *Philosophical Perspectives* 19:153–178.

——— (2010). A Defence of Imprecise Credences in Inference and Decision Making. *Philosophical Perspectives* 24:281–323.

Karni, Edi and Schmeidler, David (2016). An expected utility theory for state-dependent preferences. *Theory and Decision* 81:467–478.

Karni, Edi and Vierø, Marie-Louise (2013). "Reverse Bayesianism": A Choice-Based Theory of Growing Awareness. *American Economic Review* 103(7):2790–2810.

Keynes, John Maynard (1921). *Treatise on Probability*. London: Macmillan & Co.

Konek, Jason (forthcoming). Epistemic Conservativity and Imprecise Credence. *Philosophy and Phenomenological Research*.

Koopman, Bernard O. (1940a). The Axioms and Algebra of Intuitive Probability. *Annals of Mathematics* 41(2):262–292.

——— (1940b). The Bases of Probability. *Bulletin for the American Mathematical Society* 46:763–774.

Kraft, Charles H., Pratt, John W. and Seidenberg, A. (1959). Intuitive Probability on Finite Sets. *Annals of Mathematical Statistics* 30(2):408–419.

Krantz, David H., R. Duncan Luce, Patrick Suppes and Amos Tversky (1971). *Foundations of Measurement, Vol 1: Additive and Polynomial Representations*. Academic Press.

Kripke, Saul A. (1982). *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.

Leitgeb, Hannes (2017). *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford: Oxford University Press.

Levi, Isaac (1980). *The Enterprise of Knowledge*, Cambridge, Mass.: MIT Press.

Lewis, David (1980). A Subjectivist's Guide to Objective Chance. In Jeffrey, Richard C. (*ed.*), *Studies in Inductive Logic and Probability*, Volume II, Berkeley: University of California Press:263–293.

List, Christian and Pivato, Marcus (2015). Emergent Chance. *The Philosophical Review* 124(1):119–152.

Lockhart, Ted (2000). *Moral Uncertainty and its Consequences*. Oxford University Press.

MacAskill, William (2014). *Normative Uncertainty*. DPhil thesis, Oxford University.

——— (2016a). Smokers, Psychos, and Decision-Theoretic Uncertainty. *The Journal of Philosophy* 113(9):425–445.

——— (2016b). Normative Uncertainty as a Voting Problem. *Mind* 125(500):967–1004.

MacAskill, William and Ord, Toby (forthcoming). Why Maximize Expected Choice-Worthiness? *Noûs*.

Maher, Patrick (1997). Depragmatized Dutch Book Arguments. *Philosophy of Science 64*(2):291–305.

Mahtani, Anna (2012). Diachronic Dutch Book Arguments. *The Philosophical Review* 121(3):443–450.

——— (2015). Dutch Books, Coherence, and Logical Consistency. *Noûs* 49(3):522–537.

——— (2018). Imprecise Probabilities and Unstable Betting Behaviour. *Noûs* 52 (1):69-87.

Mayo-Wilson, Conor and Wheeler, Gregory (2016). Scoring Imprecise Credences: A Mildly Immodest Proposal. *Philosophy and Phenomenological Research* 92(1):55-78

Meacham, Christopher J. G. and Weisberg, Jonathan (2011). Representation Theorems and the Foundations of Decision Theory. *Australasian Journal of Philosophy* 89(4):641–663.

Moller, Dan (2011). Abortion and Moral Risk. *Philosophy* 86(3):425–443.

Nissan-Rozen, Ittay (2015). A Triviality Result for the "Desire by Necessity" Thesis. *Synthese* 192(8):2535–2556.

Olsson, Erik J. (2005). *Against Coherence: Truth, Probability, and Justification.* Oxford: Oxford University Press.

Pettigrew, Richard (2013). A New Epistemic Utility Argument for the Principal Principle. *Episteme* 10(1):19–35.

——— (2016a). *Accuracy and the Laws of Credence*. Oxford University Press.

——— (2016b). Accuracy, Risk, and the Principle of Indifference. *Philosophy and Phenomenological Research* 92(1):35–59.

——— (2016c). Epistemic Utility Arguments for Probabilism. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.

Predd, Joel B., Seiringer, Robert, Lieb, Elliott H., Osherson, Daniel N., Poor, H. Vincent, and Kulkarni, Sanjeev R. (2009). Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions on Information Theory* 55(10):4786–4792.

Ramsey, Frank P. (1926). Truth and probability. In Richard B. Braithwaite (ed.) *The Foundations of Mathematics and other Logical Essays*. Routledge and Kegan Paul.

Rawls, John (1951). Outline of a Decision Procedure for Ethics. *The Philosophical Review* 60(2):177–197.

——— (1975). The Independence of Moral Theory. *Proceedings and Addresses of the American Philosophical Association.*

——— (1999). *A Theory of Justice*, Revised Edition. Cambridge, MA: Harvard University Press.

Riedener, Stefan (2015). *Maximising Expected Value Under Axiological Uncertainty: An Axiomatic Approach.* DPhil thesis, Oxford University.

Rinard, Susanna (2013). Against Radical Credal Imprecision. *Thought: A Journal of Philosophy* 2(1):157–165.

Rosenkrantz, Roger (1981). *Foundations and Applications of Inductive Probability.* Atascadero, CA: Ridgeview Press.

Ross, Jacob (2006). *Acceptance and Practical Reason*. PhD thesis, Rutgers University.

Savage, Leonard J. (1954). *The Foundations of Statistics*. Wiley Publications in Statistics.

Scanlon, T.M. (2003). Rawls on Justification. In Samuel Freeman (ed.), *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press.

Schaffer, Jonathan (2007). Deterministic chance?. *British Journal for the Philosophy of Science* 5(2):113–140.

Schipper, Burkhard C. (2014). Awareness. In Hans van Ditmarsch, Joseph Y. Halpern, Wiebe van der Hoek and Barteld Kooi (eds.) *Handbook of Epistemic Logic*. College Publications.

Schoenfield, Miriam (2017a). The Accuracy and Rationality of Imprecise Credences. *Noûs* 51(4):667–685.

——— (2017b). Conditionalization Does Not Maximize Expected Accuracy. *Mind* 126(504):1155–1187.

Scott, Dana (1964). Measurement structures and linear inequalities. *Journal of Mathematical Psychology* 1:233–247.

Schroeder, Mark (2008). *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.

Schwitzgebel, Eric (2010). Kant on Killing Bastards, on Masturbation, on Wives and Servants, On Organ Donation, Homosexuality, and Tyrants. *The Splintered Mind*. `http://schwitzsplinters.blogspot.com/2010/03/kant-on-killing-bastards-on.html`

Seidenfeld, Teddy, Schervish, Mark J. and Kadane, Joseph B. (2012). Forecasting with imprecise probabilities. *International Journal of Approximate Reasoning* 53:1248–1261.

Seidenfeld, Teddy and Wasserman, Larry (1993). Dilation for Sets of Probabilities. *The Annals of Statistics* 21(3):1139–1154.

Sepielli, Andrew (2009). What to Do When You Don't Know What to Do. *Oxford Studies in Metaethics* 4:5–28.

——— (2010). *Along an Imperfectly-Lighted Path*. PhD thesis, Rutgers University. `https://rucore.libraries.rutgers.edu/rutgers-lib/26567/`

——— (2011). Normative Uncertainty for Non-Cognitivists. *Philosophical Studies* 160(2):191–207.

——— (2012). Subjective Normativity and Action Guidance. *Oxford Studies in Normative Ethics* 2.

——— (2013). What to Do When You Don't Know What to Do When You Don't Know What to Do... *Noûs* 47(1):521–544.

Shimony, Abner (1955). Coherence and the Axioms of Confirmation. *Journal of Symbolic Logic* 20:1–28.

——— (1970). Scientific Inference. In Robert G. Colodny (ed.), *The Nature and Function of Scientific Theories.* Pittsburgh: University of Pittsburgh Press.

Singer, Peter (1972). Famine, Affluence, and Morality. *Philosophy Public Affairs* 1:229–243.

Skyrms, Brian (1995). Strict Coherence, Sigma Coherence, and the Metaphysics of Quantity. *Philosophical Studies* 77(1):39–55.

Smith, Michael (2002). Evaluation, Uncertainty and Motivation. *Ethical Theory and Moral Practice* 5(3):35–320.

Staffel, Julia (forthcoming). Expressivism, Normative Uncertainty, and Arguments for Probabilism. *Oxford Studies in Epistemology*.

Stefánsson, H. Orri and Steele, Katie (manuscript). Belief Revision for Growing Awareness.

Strevens, Michael (1998). Inferring Probabilities from Symmetries. *Noûs* 32(2):231–246.

Suppes, Patrick (1969). *Studies in the Methodology and Foundations of Science: Selected Papers from 1951 to 1969*. Springer.

Teller, Paul (1973). Conditionalization and observation. *Synthese* 26(2):218–258.

Thomson, Judith Jarvis (1971). A Defense of Abortion. *Philosophy & Public Affairs* 1:47–66.

Titelbaum, Michael G. (manuscript). *Fundamentals of Bayesian Epistemology*.

Vallinder, Aron (2018). Imprecise Bayesianism and Global Belief Inertia. *The British Journal for the Philosophy of Science* 69(4):1205-1230.

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

——— (1989). *Laws and Symmetry*, Oxford: Clarendon Press.

——— (1990). Figures in a Probability Landscape. In Dunn, J. M. and Gupta, A. (*eds.*), *Truth or Consequences: Essays in Honor of Nuel Belnap*, Dordrecth: Kluwer.

——— (1995). Belief and the Problem of Ulysses and the Sirens. *Philosophical Studies* 77:7–37.

Villegas, Carlos (1964). On qualitative probability σ-algebras. *Annals of Mathematical Statistics* 35(4):1787–1796.

Vineberg, Susan (2016). Dutch Book Arguments. In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition).

von Neumann, John and Morgenstern, Oskar (1944/2007). *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.

Walley, Peter (1991). *Statistical Reasoning with Imprecise Probabilities*. Monographs on Statistics and Applied Probability, Vol 42. London: Chapman and Hall.

Wittgenstein, Ludwig (1922). *Tractatus Logico-Philosophicus.* Translated by C.K. Ogden. London: Kegan Paul, Trench, Trubner & Co.

——— (1953/2009). *Philosophical Investigations.* Translated by G.E.M. Anscombe, P.M.S. Hacker, and Joachim Schulte. Revised fourth edition by P.M.S. Hacker and Joachim Schulte. Chichester: Wiley-Blackwell.

Weatherson, Brian (2014). Running Risks Morally. *Philosophical Studies* 167(1):141–163.

White, Roger (2005). Epistemic Permissiveness. *Philosophical Perspectives.* 19:445–459.

——— (2010). Evidential Symmetry and Mushy Credence. *Oxford Studies in Epistemology* 3:161–186.

Williamson, Jon (2010). *In Defence of Objective Bayesianism*. Oxford: Oxford University Press.

Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.