Issues in accuracy epistemology

Matthew Hewson

Merton College

University of Oxford

A thesis submitted for the degree of DPhil in Philosophy

67,606 words

April 2022

To my mother, my father, and my wife.

I know Fox News is the target of criticism but we take it in stride. Our number one goal is to be accurate.

— Shannon Bream

Preface

I'm not sure when I first came across accuracy epistemology but it must have been roughly 2014, when Richard Pettigrew gave a talk at the University of Birmingham — likely at the invitation of Scott Sturgeon. Working with Tim Williamson on the BPhil a couple of years later, and interested in norms on gathering evidence, he suggested I look at a paper by Bernhard Salow and Catrin Campbell-Moore, on the relationship between evidence-gathering, accuracy and epistemic risk. I found getting to grips with their paper very difficult, but immensely gratifying. Around the same time I attended a workshop on accuracy epistemology at KCL. This cemented a nascent interest. I decided to write my BPhil thesis on the philosophical foundations of the programme. I've yet to get bored.

Working on this thesis has been an enormous privilege. The last couple of years have been extremely challenging, for the obvious reasons and some more besides. I've tried very hard to not lose sight of just how lucky I am to have been afforded this time to work on questions I find so interesting. It is not an opportunity many get.

I owe a lot of thanks to a lot of people.

I am extremely grateful for the financial support I received, without which, my time at Oxford would not have been remotely possible. I thank the AHRC and Pembroke College for financial support in the period 2016-18, when I completed the BPhil. I thank the Clarendon Fund and Merton College for DPhil funding, and the Aristotelian Society for a further award in the 2021-22 academic year. I am also very grateful to Merton College for funding my attendance at CoPS-FaM-19 and the 10th Edinburgh Graduate Epistemology Conference.

Tim Williamson's reputation as a philosopher is rightly legendary. He is as excellent a supervisor as he is a philosopher. It has sincerely been the privilege of my academic life to work with him, and if this thesis reflects even a fraction of the

vast amount he has taught me over the last six years, I will be very happy. Bernhard Salow was the perfect secondary supervisor: always able to see the wood for the trees; ever on-hand with solutions if I had problems; and with a relentless supply of interesting problems just when I thought I had solutions. Reading his and Catrin Campbell-Moore's work during the BPhil had a huge impact on the direction of my philosophical life, and it was both fitting and rewarding to have completed this thesis under his co-supervision.

I received incredibly helpful comments on various parts of the thesis from Kevin Dorst, Kenny Easwaran, Branden Fitelson, Pavel Janda and Daniel Rothschild. I received feedback on material from chapter one from CoPS-FaM-19 (where Ben Levinstein was a commentator); on material from chapter three at Open Minds XIII and the Oxford Epistemology Group; on material from chapter five at the 10th Edinburgh Graduate Epistemology conference (where Clare Field was a commentator). I am very thankful for all the feedback I got at these events.

I am indebted to Scott Sturgeon, who was an immense figure of support when I was at the University of Birmingham. Were it not for his help and guidance, I would not be at Oxford, and this thesis would not exist. I hope the fact that it speaks to some of the issues that he is interested in — and which he got me interested in — can serve as a small mark of the esteem in which I hold him.

I thank Umut Baysan for being a constant source of advice in college. I thank Iris Geens — the glue that holds the department together — for answering probably thousands of my emails about some BPhil/DPhil regulation or other.

Material in chapter one is derived, in large part, from Hewson (2020) and Hewson (2021). I am grateful to *Analysis* and *Noûs* for allowing me to include that material here, and to the referees and editors for those journals that shepherded these manuscripts from submission to acceptance. I am also grateful for comments from editors and referees on manuscripts, based on material from this

thesis, that were not accepted. Material in chapter three is derived (admittedly, barely recognisably) from part of my BPhil thesis, submitted in 2018.

The graduate community in Oxford, past and present, has been as reliable a source of support — intellectual, moral, practical — as I could have hoped for. In addition to receiving their valuable philosophical feedback, I have had enriching conversations in the department, restorative coffee at Common Ground, memorable evenings at the Rose & Crown, or some combination of the three, with Milena Bartholain, Michael Bruckner, Pietro Cibinel, Tomi Francis, William Gildea, AJ Gilbert, Eilidh Guthrie, James Kirkpatrick, Kacper Kowalczyk, Lukas Lewerentz, Sebastian Liu, Annina Loets, Aglaia von Lötz, Chiara Martini, Richard Roth, Beatriz Santos and last but certainly not least, Weng Kin San. Though not a philosopher, William Brockbank taught me more about philosophy than he probably knows.

I thank Charlotte Caluori and Abraham Martin-Rodriguez for introducing me to philosophy. I express my sincere gratitude for the friendship of Paul Bacchus and Kash Sunghuttee, both of whom were with me at formative parts of my philosophical education.

To my siblings: thank you for supporting me during this strange process, and always knowing when to ask — and when not to ask — how it was going.

To my parents: at not one point in the last decade have I felt that you questioned my choice to do something as obscure as study philosophy. Your support has been completely unfailing throughout my life, and I could not have done any of this without your love and care.

To Hannah: thank you. For your love, your friendship, for upping sticks to be with me here — for everything you've done to support me in this. I am astonishingly fortunate to be your partner, and before long, your husband. I'd be sad and blue, if not for you.

Abstract.

This thesis looks at various questions in the accuracy epistemology research programme. These include: what kind of accuracy coherence (if any) grounds epistemic requirements on outright belief, and what do the resulting requirements (if any) look like; how does this kind of accuracy coherence constrain or otherwise bear on higher-order beliefs; how should we reason when we are uncertain what the correct theory of accuracy (understood as the epistemic good) is; can accounts of expectedly accurate updates make sense of the idea that it is epistemically valuable to gather new evidence; can accounts of expectedly accurate updates make proper sense of how we should respond to new evidence.

Table of Contents

0 Introduction **1**

1 Accurate believers are deductively cogent **5**

1.1 Introduction 5

1.2 Two notions of accuracy coherence 7

1.2.1 W-coherence 10

1.2.2 $\cap B$-coherence 11

1.3 Consistency and closure 14

1.4 Accuracy arguments for norms on outright belief 18

1.5 Two kinds of accuracy coherence 23

1.5.1 $\cap B$-coherence and propriety 24

1.5.2 Pairwise consistency 29

1.5.3 Single-premise closure 30

1.5.4 Summary 35

1.6 Objections and replies 36

1.7 Further discussion 46

1.7.1 The relationship between belief and credence 46

1.7.2 Alternative accuracy functions 51

1.7.3 An analogy with inconsistent preferences 57

1.7.4 Updating 58

1.8 Conclusion 59

Appendix 1.A Ultra-conservative W-coherence guarantees consistency 60

Appendix 1.B W-coherence never requires multi-premise closure . . . 64

Appendix 1.C Single-premise closure is not required for maximising expected accuracy 65

2 Accuracy and higher-order belief **67**

2.1 Introduction 67

2.2 Accuracy coherence and doxastic logic frames 68

2.2.1 Formal background 68

2.2.2 $\cap$-coherence 69

2.2.3 Necessary and sufficient conditions for $\cap$-coherence of B 71

2.2.4 Doxastic logic 80

2.2.5 Accuracy Frame $\cap$-coherence 81

2.3 Belief introspection 85

2.4 Anti-expertise 88

2.4.1 Some properties of self-effacing frames 92

2.4.2 Connections between self-effacing frames and anti-expertise 96

2.4.3 Omissive anti-expertise 99

2.5 Moore propositions 101

2.6 An alternative picture 105

2.7 Conclusion 112

Appendix 2.A Accuracy frame W-coherence 112

Appendix 2.B More results about self-effacing frames 117

2.B.1 Sufficient conditions for avoiding self-effacing frames when $T > |F|$ 118

2.B.2 Sufficient conditions for avoiding self-effacing frames when $|F| \geq T$ 120

Appendix 2.C Self-effacing frames and belief introspection 125

3 Normative uncertainty and the accuracy argument for probabilism **129**

3.1 Introduction 129

3.2 The problem 132

3.3 A moral analogue 139

3.3.1 Normative Internalism 142
3.3.1.1 My Favourite Theory 142
3.3.1.2 Maximise Expected Choiceworthiness 143
3.3.2 Normative Externalism 145
3.3.3 Summary 146
3.4 The solution 147
3.4.1 My Favourite Theory 147
3.4.2 Maximise Expected Choiceworthiness 148
3.4.3 Normative Externalism 151
3.4.4 Summary 158
3.5 Further issues 158
3.5.1 Uncertainty about uncertaintist rules 158
3.5.2 Other accuracy arguments 159
3.5.3 Non-additive rules 160
3.5.4 Outright belief 161
3.6 Conclusion 162
Appendix 3.A Appendix 163

4 Should you care about the way you look? 167
4.1 Introduction 168
4.2 Background 170
4.3 Good and Oddie 175
4.4 A failure of Oddie's Theorem 184
4.4.1 Measuring accuracy 184
4.4.2 A counterexample 186
4.5 Why condition? 189
4.5.1 Evaluating updates globally 194
4.5.2 Local updates 196

4.5.3 Conditioning is accuracy-conducive (given certain constraints) . . . 201
4.6 Ensuring accurate updates . . . 207
4.7 Concluding remarks . . . 211
Appendix 4.A Appendix . . . 212

5 Some hard truths about accurate updating 215
5.1 Introduction . . . 215
5.2 Accurate updating and evidence externalism . . . 217
5.3 Three lessons about accurate updating given externalism . . . 221
5.4 The bad news . . . 234
5.5 Some unpromising ways out . . . 236
5.5.1 Why worry about asymmetry . . . 236
5.5.2 Reconceiving the 'evidence for' relation . . . 238
5.5.3 Appealing to non-idealised rationality . . . 243
5.6 A more promising way out . . . 246
5.7 Conclusion . . . 253

6 Conclusion 255
References . . . 259

0 | Introduction

It's good to believe the true, and bad to believe the false. That apparently obvious claim is the unmoved mover behind pretty much this entire thesis. Yes, we need to invoke certain decision rules; the story is complicated by degrees of belief; complicated further by theories of evidence. But at its core, this thesis investigates just how far we can get, epistemologically, by taking seriously the simple, almost trivial, claim that believing true things is good for you, and believing false things is bad for you. Not good in any way, though. In an *epistemic* way. As a basis for our discussion we assume that believing the true, and not believing the false, is not (merely) practically useful, or morally, aesthetically, and so on. Rather, it is good for you insofar as you want (or ought) to have an accurate picture of the world.

In that respect, this thesis is of a piece with much recent work in epistemology, and formal epistemology. It looks at some questions in what has become known as 'accuracy epistemology'. I will play a bit fast and loose with that terminology. Depending on exactly how one interprets the jargon, there may be some subtle differences between 'accuracy-first' epistemologies, 'accuracy-only' epistemologies, 'accuracy-centred' epistemologies, and epistemic utility theories (see Joyce (2018) for some discussion of this). For the most part, I will just lump these together, and take them to mean an epistemological theory that tries to ground all epistemic requirements, one way or the other, in the goodness of believing true things and the badness of believing false things — or what is sometimes called a *veritist* value theory (Goldman, 1999; Pettigrew, 2016a, Intro). If there's ever a point at which I should write more carefully, I will (or I hope that I will, anyway), and as we'll see at a couple of points, the difference between accuracy epistemology and epistemic utility theory (where that is made precise as we proceed) does turn out to be interesting.

The promise of the accuracy-epistemological approach is neatly encapsulated in the accuracy argument for probabilism; that is, the requirement that your degrees of belief should be representable as a probability function. As the accuracy-epistemological lore goes: arguments that your beliefs should be so constrained have standardly taken one of two forms. On the one hand, there are so-called 'Dutch Book' arguments that endeavour to show that if your degrees of belief are not representable as a probability function, you'll be committed to a set of gambles that are guaranteed to lose you money (Ramsey, 1926; de Finetti, 1937). On the other hand, there are so-called 'representation theorems' which show that, if your preferences satisfy certain constraints, you will be representable as having a probability function (Ramsey, 1926; Savage, 1954). A natural complaint about these arguments is that they only give you prudential or practical reason to be probabilistic, if they give you any such reason at all (Joyce, 1998). Yet probabilism seems like an *epistemic* requirement — a requirement that should fall out of trying to represent the world aright, not out of trying to be practically successful. Accuracy arguments aim to fill this lacuna. They show that if your degrees of belief are not representable as a probability function, then they are guaranteed to be further from the ideally accurate credences — the credences that are certain all truths are true, and all falsehoods are false — than necessary (Joyce, 1998; Leitgeb & Pettigrew, 2010b; Pettigrew, 2016a). Since approximating the ideally accurate credences is (presumably) an epistemic good, this gives an epistemic explanation for the apparent badness of failing to be probabilistic.

All of the accuracy arguments we look at in this thesis (which includes the accuracy argument for probabilism) take roughly this form, attempting to show what happens when you are trying to avoid being further than necessary from an ideally accurate state, and what epistemic constraints fall out as a result.[1]

[1]Accuracy epistemologists often talk of accurate credences being 'close to the truth', rather than 'close to the ideal, omniscient, beliefs'. I will occasionally slip into this way of talking, which,

In the first chapter, however, we look not at degrees of belief, but outright belief. We ask how the set of your outright beliefs should be configured if you want them to be as accurate as possible, and in particular, we argue that a kind of accuracy-coherence that is sensitive to your perspective is needed to generate epistemic requirements of an appropriate strength. The resulting arguments seem to show that the right kind of accuracy coherence for outright beliefs gives rise to much stronger epistemic requirements on outright belief than accuracy epistemologists have standardly thought.

Chapter two extends our investigation of the account of accuracy coherence discussed in chapter one, by turning to questions about higher-order beliefs. A tentatively attractive account of the epistemology of higher-order belief emerges.

The third chapter shifts focus from outright belief to degrees of belief. In particular, we look at the accuracy argument for probabilism, and a well known problem for it that results when you are uncertain of what the true theory of accuracy is. I argue that we should model this uncertainty as a form of normative uncertainty, and that by applying theories of normative uncertainty used in ethics, the accuracy argument for probabilism can be recovered.

Chapter four shifts focus yet again, this time to a question about evidence. Gathering evidence seems like an epistemically good thing to do. It's well known that given certain constraints, including a philosophically naive theory of evidence, gathering evidence is accuracy-conducive. It's also well known that given more philosophically sophisticated accounts of evidence, the epistemic goodness of evidence-gathering no longer holds. This might look like a reason to think that accuracy epistemologists cannot vindicate the goodness of gathering evidence. I argue that this conclusion is too quick. Depending on how we assess the accuracy of updating on new pieces of evidence, the result showing that evidence-

following Pettigrew (2016a, p. 3), Sturgeon (2019, esp. pp. 1334-35), I think is misleading, but ultimately harmless for our purposes.

gathering need not be accuracy conducive might assume you are using evidence sub-optimally. When we control for this, and assume that you gather evidence, then update on it in the most accuracy-conducive way, versions of the evidence-gathering claim can be recovered.

The fifth chapter pursues these issues further. I show that the update functions appealed to in the recovery of the evidence-gathering claims above have a host of unattractive properties. This puts pressure on the idea that there can be an accuracy-theoretic story about both the goodness of gathering evidence, and the goodness of responding to evidence properly.

Chapter six draws some morals from the foregoing, and suggests some questions for further research.

1 | Accurate believers are deductively cogent

This chapter argues that the agent concerned to have accurate (outright) beliefs will have a consistent and multi-premise closed belief set, and not a (merely) single-premise closed and (merely) pairwise consistent belief set, as has often been thought. This argument rests on the fact that we need a notion of accuracy coherence for belief that is *belief-sensitive*; sensitive to your perspective, in a way that the standard belief-insensitive notion of accuracy coherence is not. The choice of the belief-sensitive over belief-insensitive notion is defended, as is the choice of this particular belief-sensitive kind of accuracy coherence over other belief-sensitive rivals. Our conclusion is that accurate believers are deductively cogent.

1.1 Introduction

Accuracy arguments for epistemic norms are increasingly popular. But so far, those arguments have overwhelmingly been for norms upon *credences*, norms like probabilism, (versions of) conditionalisation and the principal principle. There's been far less attention paid to what accuracy arguments mean for norms on outright belief. On the face of it, this is pretty surprising. For while accuracy arguments may give us a richer understanding of why certain credal norms obtain, there's already pretty wide agreement, for example, that your credences should be probabilistic, and that you should update by conditionalisation.[1] When it comes to norms for outright belief however — consistency constraints and closure conditions — there's far less agreement, as the well-known preface and lottery cases make clear. Accuracy arguments here could do more than enrich our understanding of the various norms: they could help us work out which norms obtain in the first place.

And, in fact, they can. My claim is that, at least given the standard way of

[1] We discuss accuracy arguments for probabilism in chapter three; we discuss accuracy arguments for conditionalisation, as found in Greaves and Wallace (2005); Leitgeb and Pettigrew (2010b); Gallow (2019), in chapters four and five.

measuring the accuracy of (outright) beliefs, *accurate believers are deductively cogent*; if you care about having accurate beliefs, then those beliefs should be consistent and multi-premise closed. In the first part of the paper, I introduce two different kinds of accuracy coherence. One is pretty well known in the accuracy literature; it's discussed by Easwaran (2016) and Fitelson and Easwaran (2015), for example. This notion of accuracy coherence is *belief-insensitive* since it requires that your beliefs be accurate (in the relevant sense) at all possible worlds. The second is a notion of accuracy coherence that is *belief-sensitive*, since it requires that your beliefs be accurate (in the relevant sense) at all and only those worlds your beliefs treat as live. This is far less discussed, though as we will later see, a version of this picture can be found in Pettigrew (2017). These two notions — the belief-insensitive 'W-coherence', and the belief-sensitive '$\cap B$-coherence' — have strikingly different properties. In section 1.4, we see that the former requires mere single-premise closure, and mere pairwise consistency, whereas the latter requires not only consistency proper, but multi-premise closure — in other words, *deductive cogency*.[2]

The claim that accurate believers are deductively cogent amounts, then, to the claim that it's the latter, belief-sensitive notion of accuracy coherence that's the right one. And in section 1.5, this is what I argue. First, this notion of accuracy coherence is the outright belief analogue of a linchpin credal accuracy principle known as *propriety*, a principle relied upon in virtually all credal accuracy arguments. Reasons for rejecting this notion risk being reasons for rejecting propriety, undoing too much good work in the process. Second, the belief-sensitive notion of accuracy coherence secures pairwise consistency relative to an extremely wide range of accuracy functions, whereas the belief-insensitive notion of accuracy coherence can do so only under certain difficult-to-motivate constraints. The final

[2]The label 'deductive cogency' belongs to Isaac Levi, see e.g. Levi (1967).

and most important reason relies on the observation that single-premise closure is *so* plausible that any accuracy argument which fails to generate it is no good. I then argue that there are crucial philosophical issues with the arguments given by Easwaran (2016) and Fitelson and Easwaran (2015); issues that can be fixed, but at the cost of sacrificing single-premise closure. So properly understood, the choice between our two different kinds of accuracy coherence isn't a choice between deductive cogency on the one hand, and mere single-premise closure (plus pairwise consistency) on the other. It's a choice between deductive cogency, and no closure requirements at all.

Section 1.6 considers some other objections. I investigate whether the proposal in Dorst (2019) that accurate belief sets maximise expected accuracy can recover pairwise consistency and single-premise closure. I show that it runs into exactly the same problems that faced Easwaran (2016) and Fitelson and Easwaran (2015). I argue that there are no other notions of accuracy coherence worthy of consideration. Section 1.7 looks at some other issues raised by this alternative, belief-sensitive, kind of accuracy-coherence. The chief conclusion of this chapter is sustained: accurate believers are deductively cogent.

1.2 Two notions of accuracy coherence

In this section I introduce the two notions of accuracy coherence that we're concerned with in this paper: the belief-insensitive W-coherence, and the belief-sensitive $\cap B$-coherence. But before I can do so, we need a way of thinking about propositions and belief sets that lets us evaluate their accuracy.[3]

We let W be a finite set of possible worlds, $\{w_1 ... w_n\}$, where these worlds are mutually exclusive and jointly exhaustive. We then treat propositions p as sets of

[3]This is the standard framework, used by e.g. Easwaran (2016); Fitelson and Easwaran (2015); Dorst (2019).

worlds. So, for example, the proposition that the cat is on the mat is modelled as the set of possible worlds where the cat is on the mat. A proposition p is true at a world w just in case $w \in p$. The conjunction of some propositions, $p_1 \wedge ... \wedge p_n$ is their intersection $p \cap ... \cap p_n$. The disjunction of some propositions $p_1 \vee ... \vee p_n$ is their set-theoretic union $p_1 \cup ... \cup p_n$. The negation of a proposition $\neg p$ is $W \setminus p$. We say that p entails q when $p \subseteq q$. We then treat belief sets B as sets of propositions. A proposition p is believed according to a belief set B just in case p is in B. When neither p nor its negation are in B we say that p is suspended on by B.[4]

We're now in a position to explain how the accuracy of belief sets is measured. The accuracy functions we're interested in work in the following, simple way: they count all the true beliefs in a belief set, and give each one a 'truth' score of T; they count all the false beliefs in the belief set, and give each one a 'falsity' score of F, and then they sum those totals. Since our framework says that a proposition p is true at a world w just in case $w \in p$, for our purposes accuracy functions are defined as follows.

Definition 1 (Accuracy function for belief sets)**.** *A^b is an accuracy function for belief sets iff*

$$A^b(B, w) = |\{p \in B : w \in p\}|T + |\{p \in B : w \notin p\}|F$$

Although the function we have just defined gives us a natural way to measure the accuracy of a belief set at a world, it is by no means the only way to do so.[5] We can motivate our choice of accuracy function by appealing to certain

[4]There's almost certainly more to suspended judgment than this but we'll treat it as a simplifying modelling assumption. See Friedman (2013) for interesting discussion on the nature of suspended judgment.

[5]Thanks to Kenny Easwaran for this point. I discuss some alternative ways of designing accuracy functions later; here we follow the simple functions that the literature has concerned itself with so far. I should also note at this point that I'll assume act-state independence throughout.

constraints that, intuitively, an accuracy function should satisfy.[6] First, *extensionality*: whether believing p is epistemically good or bad at a world depends only on whether p is true at that world. This falls out of the veritist value theory that is constitutive of accuracy epistemology. Secondly, *separability*: the accuracy of a belief set at a world is just the sum of the accuracies of its constituent beliefs at a world. There is no 'holistic' benefit to certain combinations of beliefs over and above what the individual beliefs themselves contribute. The accuracy functions we look at are both extensional and separable in this sense.

At this stage we will make just one assumption about the truth and falsity scores: that $T > 0 > F$.[7] This encodes the assumption that true beliefs are epistemically good and false beliefs are epistemically bad, while neither believing a proposition nor its negation is in some sense 'epistemically neutral'.

To give a sense of how all this formalism works, we can appeal to some graphical representations. Figure 1.1 represents our set of propositions when $W = \{w_1, w_2, w_3\}$. Figure 1.2 represents a belief set defined on that set of propositions, 'B' marks the believed propositions while '$*$' marks the unbelieved propositions. The upward lines denote the subset relation.

We determine the accuracy score of $\{W, \{w_1, w_2\}\}$ at a world by counting the number of true and false propositions at that world. So:

- $A^b(\{W, \{w_1, w_2\}\}, w_1) = 2T$
- $A^b(\{W, \{w_1, w_2\}\}, w_2) = 2T$

[6] Dorst (2019, p. 184) motivates this choice of accuracy function in this way. As he effectively points out, this also has the virtue of making these accuracy functions for outright belief 'axiomatically' very similar to the accuracy functions standardly used in the accuracy epistemology of credences, for these also satsify (credal analogues of) extensionality and separability. See e.g. Joyce (1998); Predd et al. (2009); Leitgeb and Pettigrew (2010a); Briggs and Pettigrew (2020).

[7] For the majority of this chapter I'll assume that all propositions are weighted equally; that the value of one being true is the same as the value of any other being true, and likewise for the disvalue of falsity. See Easwaran (2016, pp. 849-52) for more on this; I also discuss this assumption further in section 1.7.2.

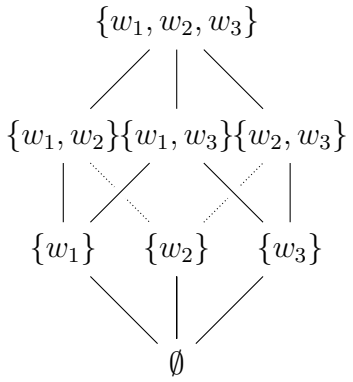


Figure 1.1: The set of propositions in our toy model

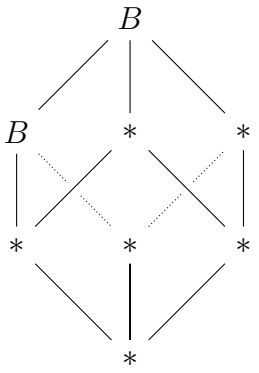


Figure 1.2: A toy belief set, where only W and $\{w_1, w_2\}$ are believed, and all other propositions are not believed.

- $A^b(\{W, \{w_1, w_2\}\}, w_3) = T + F$

With a picture of propositions and belief sets in place, we can now introduce our two notions of accuracy coherence.

1.2.1 W-coherence

The first notion of accuracy coherence for outright belief that we'll be interested in is the *belief-insensitive* notion described above (I'll say more about these labels

after I've introduced both types of coherence).[8] First, let's specify the following dominance conditions:

Definition 2 (Strong W-dominance)**.** *A belief set B' strongly W-dominates another belief set B iff $A^b(B', w) > A^b(B, w)$ at every $w \in W$.*

Definition 3 (Weak W-dominance)**.** *A belief set B' weakly W-dominates another belief set B iff $A^b(B', w) \geq A^b(B, w)$ at every $w \in W$, and moreover, $A^b(B', w) > A^b(B, w)$ for at least some $w \in W$.*

I will say that B is weakly W-dominated iff there is some B' such that B' weakly W-dominates B. I will say that B is strongly W-dominated iff there is some B' such that B' strongly W-dominates B. Anything that is strongly W-dominated is also weakly W-dominated. So far we've not made any normative claims. But that is what the following notion of accuracy coherence amounts to:

> **W-coherence:** It's rationally prohibited for B to be either strongly or weakly W-dominated.

W-coherence is therefore the requirement that no belief set should at least as well as your own at *all* worlds and better at some (Easwaran, 2016, p. 823). I'll have more to say about exactly how this notion of accuracy coherence should be motivated later in the paper. For now, I take it that there's *something* appealing about the idea that if another belief set is more accurate than your own at *every* world, then epistemically, something's gone wrong.

1.2.2 $\cap B$-coherence

This second notion of accuracy coherence for outright belief is importantly different from W-coherence. Since in our framework, any belief set B is a set of

[8] See Easwaran (2016) and Fitelson and Easwaran (2015) for this kind of accuracy coherence.

propositions, and propositions are sets of worlds, $\cap B$ is the set of worlds common to all propositions in B, or the set of worlds at which all the propositions in B can be true. In other words, it is the set of worlds that B treats as live.[9] We can use $\cap B$ to specify some different accuracy dominance conditions.

Definition 4 (Strong $\cap B$-dominance)**.** *B' strongly $\cap B$-dominates B iff $A^b(B', w) > A^b(B, w)$ at all $w \in \cap B$.*

Definition 5 (Weak $\cap B$-dominance)**.** *B' weakly $\cap B$-dominates B iff $A^b(B', w) \geq A^b(B, w)$ at all $w \in \cap B$, and $A^b(B', w) > A^b(B, w)$ at some $w \in \cap B$.*

I will say that B is weakly $\cap B$-dominated if there is some B' such that B' weakly $\cap B$-dominates B. I will say that B is strongly $\cap B$-dominated if there is some B' which strongly $\cap B$-dominates B. Again, so far we've not made any normative claim. We do so now:

> **$\cap B$-coherence:** It's rationally prohibited for B to be either strongly or weakly $\cap B$-dominated.

By evaluating B at all worlds in $\cap B$, we evaluate the belief set at all worlds left open by our beliefs. What $\cap B$-coherence captures is the thought that accurate believers should care about striking the best possible balance between truths and falsehoods at those worlds doxastically open to them. So construed, $\cap B$-coherence amounts to a sort of internal coherence constraint. Rational agents had better make sure that they are doing as well as possible at those worlds they leave open; they needn't care about the worlds their beliefs have ruled out. This is all I'll say in defence of $\cap B$-coherence for the time being, though we'll meet some more substantive arguments later on.

[9] We'll consider some other proposals for the worlds a belief set treats as live in section 1.6. These dominance conditions are discussed in Pettigrew (2017); see section 1.4 for more on the background to this conception of belief coherence.

Three things before we move on. First, we can get a feel for how these coherence notions can be used in accuracy arguments for epistemic requirements by noting that both require you to always believe the tautologous proposition. For suppose that you didn't. Then there would be some other belief set, identical to your own except for its inclusion of the tautologous proposition. And since the tautologous proposition is true at every world, it gets a score of T at every world. So this rival belief set is guaranteed to score better than your current belief set at every world in W, and since $\cap B \subseteq W$, every world in $\cap B$, too. So if you don't believe the tautology, you're neither W-coherent nor $\cap B$-coherent.[10] This example is doubly instructive, for it reminds us that both W-coherence and $\cap B$-coherence are *idealised* conceptions of epistemic rationality. Ordinary agents like you and I are presumably not epistemically required to believe a given logical truth. But both W-coherence and $\cap B$-coherence do require this.

Second, we can now properly fill out the 'belief-sensitive' and 'belief-insensitive' talk that I've been using. We say that $\cap B$-coherence is *belief-sensitive* because whether some belief set B is $\cap B$-coherent depends on facts about which worlds B treats as live. On the other hand we say that W-coherence is *belief-insensitive* because whether B is W-coherent depends entirely on B's accuracy score at *every* possible world, even if B rules some of those worlds out.

Thirdly, we can now make clear how our investigation differs from earlier versions of the accuracy project. The approach we're interested in here has pretty direct descendants in the 'cognitive' utility-theoretic work of Levi (1961, 1965, 1967) and Hempel (1962).[11] Levi was — to the best of my knowledge — the first

[10] A similar argument shows that, on pain of violating W-coherence and $\cap B$-coherence, you should never believe the contradictory proposition.

[11] A slightly less direct, though no less important, philosophical ancestor of our approach is James (1897), and the famous injunction to 'seek truth' and 'shun error' — an injunction that, as we discuss in section 1.5.2, is something we reflect in the values we give to A^b. See Pettigrew (2016b) for an explicitly epistemic utility-theoretic investigation of James' injunction.

to consider how epistemic utilities could bear on the question of when to accept a scientific hypothesis, or as he informatively puts it, "the possibilities of extending the applicability of decision theory from the domain of inductive behavior [...] into the domain of inductive inference" (Levi, 1961, p. 614). I mentioned earlier that Levi coined the expression 'deductive cogency'. He was also the first — again, to the best of my knowledge — to use the term 'epistemic utility' in the sense which is now familiar to formal epistemologists (Levi, 1961, p. 621).

Hempel also discusses epistemic utility-theoretic approaches to *acceptance* (of scientific theories), and is the first to suggest that with appropriate constraints on those utilities, you might maximise expected epistemic utility by accepting those theories of which you are at least 0.5 sure (Hempel, 1962, p. 155). Specifically, he 'tentatively' suggests that, in the notation of this chapter, $T = |F|$ (Hempel, 1962, pp. 154-55).

Our programme, by contrast, differs in a few respects. First, we are concerned with (outright) *belief*, not acceptance. Second, unlike these alternative programmes, we don't attempt to directly encode other epistemic virtues besides accuracy in the 'utility' function. I say more about this in section 1.7.2. Third, Hempel and Levi were not attempting to derive the fundamental norms themselves from epistemic utility, so constraints like deductive cogency were already imposed upon the admissible belief sets (Hempel, 1962, pp. 150-51; Levi, 1967, ch. 2). By contrast, our aim is to see which norms fall out of certain comparisons of different belief sets and their accuracy.

1.3 Consistency and closure

We have two notions of accuracy coherence. One requires that you not to be outscored at every world; the other requires that you not be outscored just at those

worlds your belief set treats as live. This subtle difference turns out to have surprising implications for the strength of norms that fall out of the respective kinds of accuracy coherence. This section gets clear on those different norms and their respective strengths. Having done so, we'll be able to show how the different kinds of accuracy coherence generate (or fail to generate) those norms.

The first properties of belief sets that we're interested in are *consistency* constraints. A belief set is consistent iff all the propositions in it can be true at once; i.e. if there's some world common to every proposition in that set. A belief set is *pairwise* consistent iff for any two propositions in that set, there's some world common to them both. Note that pairwise consistency is weaker than consistency. The set $\{\{w_1, w_2\}, \{w_2, w_3\}, \{w_3, w_1\}\}$ is pairwise consistent: take any two propositions in the set and there is some world common to them both. But it's not consistent, for there is no world common to every proposition in the set.

Next, some closure conditions. A belief set B is *single-premise closed* just in case, for any proposition p in B, all of the logical consequences of p are in B, too.[12] Single-premise closure is something like closure under disjunction. A belief set is multi-premise closed iff, for any collection of propositions in your belief set, all the consequences of their intersection are in the belief set, too.[13] Multi-premise closure is something like closure under conjunction.

From a normative point of view, pairwise consistency and single-premise closure are non-negotiable as conditions of epistemic rationality. Pairwise consistency amounts to the weakest grade of consistency (stronger grades would be triplewise consistency, quadruplewise etc.). If your belief set is not even pairwise consistent then there are two propositions in your belief set which have no world in common and so cannot both be true at once: that's to say, they are manifestly

[12]Formally: B is single-premise closed iff, if $p \in B$ and $p \subseteq q$, then $q \in B$.

[13]Formally, B is multi-premise closed iff, if $\cap B \subseteq p$ then $p \in B$. Equivalently, B is multi-premise closed iff B contains all and only supersets of $\cap B$.

inconsistent. But epistemic rationality cannot tolerate manifest inconsistencies in your belief set. So any notion of accuracy coherence worth its salt must be able to generate pairwise consistency.

It must also be able to generate single-premise closure, a highly desirable requirement that's about as plausible as probabilism in the case of credence.[14] Suppose you believe Oxford is in the UK, but you disbelieve, or remain neutral on whether it's in Europe. That is not an epistemically good state to be in. After all, the truth of 'Oxford is in the UK' suffices for the truth of 'Oxford is in Europe' — and you are committed to the truth of 'Oxford is in the UK'. The inadequacy of this epistemic state is readily explained by the requirement of single-premise closure.

Single-premise closure is also especially plausible in the sort of idealised context we are considering. This point about plausibility will be important later on, so I'll labour it. We can look at the best challenges to single-premise closure and note that they don't apply in our idealised setting.

First, we might object that *single-premise closure is too computationally demanding*. Suppose I have a belief. Should my beliefs really have to be closed under arbitrary disjunction of that belief? For normal, human agents, it's very tempting to think that they don't. We lack the necessary cognitive resources. But that isn't the case for our ideal agents; we idealise away any cognitive limitations to make our investigation tractable. So this doesn't tell against single-premise closure here.

Second, we might object because of certain sceptical scenarios, which can pro-

[14]The discussion over closure is long and storied, and complicated further by versions of the closure principle that concern knowledge rather than rational or justified belief. See Feldman (1995); Hawthorne (2003); Pryor (2004); Hawthorne (2005) among many others for sympathetic discussion of single-premise closure (or extremely similar principles). Note that knowledge-theoretic versions of the principle are arguably harder to defend than rational-belief-theoretic ones; assuming as much, the good standing of the former further supports the standing of the latter.

vide putative cases of single-premise closure failure. It's rational for me to believe that the animal in the zoo enclosure is a zebra. By single-premise closure it's rational for me to believe it's not a cleverly-disguised horse. But (so it seems) it's hard to rule out the falsity of the latter. There's a wide literature on these kinds of problems (the consensus of which, incidentally, is probably mostly that we should hold on to single-premise closure) but in any case, in our setting, single-premise closure is *content-neutral*; it expresses a general structural requirement and abstracts away from any particular issues to do with the content of the relevant beliefs. Here is another way of putting this: if the instances of single-premise closure failure we find later on have nothing whatsoever to do with sceptical scenarios, then we can set aside possible sceptical problems for single-premise closure.

Another reason for endorsing single-premise closure is that it falls out of the two chief 'credence-first' accounts of the relation of belief to credence. *Lockeanism* holds that you believe p iff your credence in p is greater than t (where t is in the unit interval $[0, 1]$).[15] *Belief-as-credence-1* holds that you believe p iff your credence in p is 1.[16] Both views secure single-premise closure, for by finite additivity, if $p \subseteq q$, then $Pr(p) \leq Pr(q)$. And so if $Pr(p) \geq t$, or $Pr(p) = 1$ (as required by the respective views), then $Pr(q) \geq t$ or $Pr(p) = 1$.[17] In summary, single-premise closure is an exceedingly attractive principle of epistemic rationality; any notion of accuracy coherence worth taking seriously should require that your beliefs are single-premise closed.

Whether consistency and multi-premise closure are desirable is much more contentious, as the lottery and preface paradoxes famously advertise.[18] If consis-

[15] See Sturgeon (2008); Foley (2009); Dorst (2019) for defence of Lockeanism.

[16] At least insofar as your credence function is defined over a finite algebra. See Clarke (2013); Moss (2018); Dorst (2019) for defence of the belief-as-credence 1 view.

[17] And of course, because belief-as-credence-1 is just the special case of Lockeanism where $t = 1$. The principle that if $p \in B$ and $Pr(q) \geq Pr(p)$, then $q \in B$ is sometimes called 'belief monotonicity'.

[18] See Kyburg (1961) for the lottery paradox, Makinson (1965) for the preface. The literature

tency is required, then you cannot believe of each ticket that it is a loser, while also believing that one ticket will win. If multi-premise closure is required, you must believe not only of each claim in the book that it is true, but that their conjunction is true, too — and even pairwise consistency precludes both believing this and the preface-writer's claim that at least one of the claims in the book is false.

Pairwise consistency and single-premise closure are so desirable that any notion of accuracy coherence which cannot generate them is too weak to be a genuine notion of belief coherence. Consistency and multi-premise closure, though, are norms the adoption of which can be guided by accuracy arguments: it's unclear whether they hold or not, and accuracy arguments can be of use here in deciding to whether to endorse them.

1.4 Accuracy arguments for norms on outright belief

We can now start to investigate which norms fall out of our respective notions of accuracy coherence. First, W-coherence guarantees both pairwise consistency and single-premise closure. As we saw above, these are the minimal desiderata that any worthwhile notion of accuracy coherence has to satisfy.

Proposition 6. *If $|F| > T > 0$, then if B is not pairwise consistent, B is strongly W-dominated.*

Proof. Suppose B is not pairwise consistent. Then there are $p, q \in B$ such that $p \neq q$ and $p \cap q = \emptyset$. Since $p \cap q = \emptyset$, then at each world in W, at least one of p and q is false. But since $|F| + T < 0$, then suspending on both p and q is

generated by these problems is vast, see Kaplan (1996); Christensen (2004) for book-length treatments that remain highly influential.

guaranteed to be at better than believing p and q at every world. So B is strongly W-dominated (Easwaran, 2016, p. 824).[19] □

Proposition 7. *If B is not single-premise closed then B is weakly W-dominated.*

Proof. To see this, suppose B is not single-premise closed, so there is some p, q such that $p \in B, p \subset q$ and $q \notin B$. Let B' be identical to B except that it swaps p for q: $B' = B \setminus \{p\} \cup \{q\}$. Since $p \subset q$, any world in W is in exactly one of the following sets: the set of worlds in p (and therefore q), the set of worlds not in q (and therefore not in p), and the set of worlds in q, but not p.

At the worlds in p, $A^b(B, w) = A^b(B', w)$. Likewise, at those worlds not in q, $A^b(B, w) = A^b(B', w)$. However, at those worlds q but not in p, $A^b(B, w) < A^b(B', w)$, since $A^b(B', w) = A^b(B, w) + T - F$. Every world is either in q or $W \setminus q$, so there are no cases left to consider. Therefore B' weakly W-dominates B, for by definition of weak W-dominance it scores at least as well as B at all $w \in W$, and strictly better than B at some (Easwaran, 2016, p. 825). □

$\cap B$-coherence, on the other hand, is much more demanding. First, it requires consistency:

Proposition 8. *If B is inconsistent, then B is strongly $\cap B$-dominated.*

Proof. Suppose B is inconsistent. Then $\cap B = \emptyset$, for there is no world common to every proposition in B. The strong $\cap B$-dominance condition says that B' dominates B iff $A^b(B', w) > A^b(B, w)$ at all $w \in \cap B$. Since $\cap B = \emptyset$, the universally quantified statement is vacuously true and B is strongly $\cap B$-dominated — by all belief sets, including itself.[20] □

[19] It's a good question why the values of T and F are restricted in the way we've done so here; we'll come back to this later on.

[20] That $\cap B$-coherence secures consistency as a result of vacuous satisfaction of the universal quantifier might make one worry that the accuracy argument is *trivial*, and therefore lacks force. I claim: not so. See section 1.6, below, for more. This vacuity feature of $\cap B$-(in)coherence has

That $\cap B$-coherence (or something like it) requires consistency has been noted before, by Pettigrew (2017, p. 479). I say 'something like' because as best as I can tell, Pettigrew takes *something like* $\cap B$-coherence to make consistency necessary *and* sufficient. He gives 'doxastic' dominance conditions akin to the $\cap B$ ones described above. Then, he writes:

> Pick a set of consistent beliefs and disbeliefs. These then determine a set of doxastically possible worlds. At each of these worlds, each of the beliefs you picked is true and each of the disbeliefs you picked is false. Thus, you have maximal epistemic utility at each of these worlds. Any alternative assignment of beliefs, suspensions, and disbeliefs to the same propositions will be weakly doxastically dominated. Thus, any consistent set of beliefs renders itself the only rational option. (Pettigrew, 2017, p. 479)

Pettigrew is working in a very slightly different setting where one has sets of disbeliefs in addition to beliefs. These differences are inessential. It is not true of the $\cap B$-dominance conditions we use that any consistent set of beliefs has maximal epistemic utility at all of the worlds in $\cap B$. For a simple example, let $W = \{w_1, w_2\}$ and $B = \{\{w_1\}\}$. B is consistent, but it is strongly $\cap B$-dominated by $\{W, \{w_1\}\}$.

We can also show that $\cap B$-coherence requires multi-premise closure. To prove this we first need a lemma that says, if B is consistent, then at those worlds in $\cap B$, the accuracy score of B is just T multiplied by the number of propositions in B.

some other notable consequences. First, there are belief sets which are strongly $\cap B$-dominated but not weakly $\cap B$-dominated. By proposition 8, whenever B is inconsistent, it is strongly $\cap B$-dominated. However B is not weakly $\cap B$-dominated, for by its inconsistency there is no $w \in \cap B$ at which any B' outscores B. Second, though, if B is consistent, then if B is strongly $\cap B$-dominated, it is weakly $\cap B$-dominated. For suppose B is consistent and strongly $\cap B$-dominated. Since B is consistent then $\cap B \neq \emptyset$. Since B is strongly dominated, there is some B' such that $A^b(B, w) < A^b(B', w)$ at all $w \in \cap B$. It follows that $A^b(B, w) \leq A^b(B', w)$ at all $w \in \cap B$ and $A^b(B, w) < A^b(B', w)$ at some $w \in \cap B$, which just is the definition of weak $\cap B$-dominance.

Lemma 9. *If B is consistent, then at all $w \in \cap B$, $A^b(B, w) = |B|T$.*

Proof. Since B is consistent, all $p \in B$ are true at those $w \in \cap B$ (for $\cap B$ is just the set of worlds common to every proposition in B). By the definition for accuracy functions: $A^b(B, w) = |\{p \in B : w \in p\}|T + |\{p \in B : w \notin p\}|F$. But when $w \in \cap B$, there is no $p \in B$ such that $w \notin p$. Therefore when B is consistent and $w \in \cap B$: $A^b(B, w) = |\{p \in B : w \in p\}|T$. And since every $p \in B$ is true at $w \in \cap B$, then whenever $w \in \cap B$, it follows that $A^b(B, w) = |B|T$. □

Proposition 10. *If B is not multi-premise closed, then B is strongly $\cap B$-dominated.*

Proof. Suppose B is inconsistent, and not multi-premise closed. Then by proposition 8, B is strongly $\cap B$-dominated. Next, suppose B is consistent, but not multi-premise closed. Then by definition of multi-premise closure, there is some p such that $\cap B \subseteq p$ and $p \notin B$. Therefore by lemma 9 $A^b(B \cup \{p\}, w) > A^b(B, w)$ at all $w \in \cap B$. □

The crucial part of the above is the following: if your belief set B is consistent but not multi-premise closed, then it lacks some proposition p true at all the worlds in $\cap B$. But then another belief set, identical to your own but for its inclusion of p is guaranteed to do better at all worlds in $\cap B$.

So $\cap B$-coherence requires consistency and multi-premise closure, or as that combination is also known, *deductive cogency.* Since consistency entails pairwise consistency, and multi-premise closure entails single-premise closure, it follows that both W-coherence and $\cap B$-coherence are able to secure these non-negotiable epistemic requirements.

Given that both kinds of accuracy coherence guarantee single-premise closure and pairwise consistency, it's worth checking that the two notions actually come apart. That is: it's worth checking that W-coherence doesn't require consistency and multi-premise closure, too.

To see that the two notions can come apart, we can consider a particular toy belief set. The belief set depicted in Figure 1.3 is pairwise consistent, single-premise closed, but inconsistent and multi-premise closed. Moreover, at least for some values of T and F, it's W-coherent.[21] But since it's not consistent, and not multi-premise closed, it's not $\cap B$-coherent: $\cap B$-coherence and W-coherence are different normative requirements.

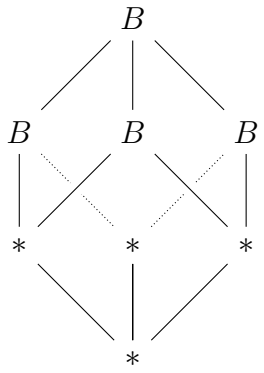


Figure 1.3: When $T < |F| < 2T$, this belief set is merely single-premise closed, and merely pairwise consistent, but nonetheless, W-coherent.

We might still worry that this difference depends on the values we've given to T and F. For altering those values can bring the two kinds of coherence closer together: if $|F| > (|W| - 1)T$, then W-coherence requires consistency after all.[22] But it turns out there are no accuracy functions (values of T and F) whatsoever that mean W-coherence requires multi-premise closure.[23] So W-coherence *is* a genuinely weaker accuracy-theoretic requirement than $\cap B$-coherence.[24]

This means our two notions of accuracy coherence are not just informally, but

[21] See Easwaran (2016, p. 829) for a proof.
[22] See Appendix 1.A for a proof.
[23] See Appendix 1.B for a proof.
[24] Weaker in the sense of the closure requirements it generates; the logical relationship between the two sets of dominance conditions is more complicated, as proposition 115 shows, and as we discuss in section 1.6.

formally different. What follows, given our starting point of using accuracy arguments to work out what norms there are for outright belief? I claim, it leaves us at stalemate, between those who wish to use accuracy-theoretic reasoning to endorse deductive cogency, and those who wish to use the same reasoning to reject it. Given the state of play in the literature, this stalemate is arguably better for proponents of deductive cogency. Fitelson and Easwaran (2015) have already given accuracy arguments for e.g. inconsistency-permitting treatments of the lottery and preface. Now defenders of deductive cogency can do likewise, the accuracy-centred advantage enjoyed by the first kind of view has been lost. But we needn't rest at stalemate. In the next section I show that there are good reasons to think $\cap B$-coherence is the notion of accuracy coherence we should endorse.

1.5 Two kinds of accuracy coherence

We have two different accuracy-theoretic notions of coherence, and two corresponding accuracy arguments. The former gives us *mere* single-premise closure (and *mere* pairwise consistency). The latter gives us deductive cogency. How should we adjudicate?

That is my task in this section. I want to raise three issues, all of which tell in favour of $\cap B$-coherence rather than W-coherence. The first concerns a structural analogy between $\cap B$-coherence and a principle extensively relied upon in giving credal accuracy arguments, known as *propriety*. The second is that $\cap B$-coherence secures pairwise consistency relative to a very broad class of accuracy functions, whereas W-coherence only secures pairwise consistency relative to a much more limited class of functions, the restrictions on which are not easy to justify. The third is that the fact W-coherence can secure single-premise closure at all relies on an apparent conflation: between the set of worlds used to generate a body

of propositions, and the set of worlds that are epistemically open. Fixing the conflation means W-coherence no longer requires single-premise closure. Let's take these in turn.

1.5.1 $\cap B$-coherence and propriety

$\cap B$-coherence looks something like the outright belief analogue of a credal condition called *propriety*. To explain what propriety says, we need some new resources.

Definition 11 (Credence function)**.** *A credence function Cr is a function $\mathcal{P}(W) \rightarrow [0, 1]$.*

Definition 12 (Probability function)**.** *A probability function Pr is a credence function Cr that satisfies the following requirements:*

$$Pr(W) = 1 \qquad \text{[Maximality]}$$

$$Pr(p) \geq 0 \qquad \text{[Non-negativity]}$$

$$Pr(p \cup q) = Pr(p) + Pr(q) \textit{ whenever } p \cap q = \emptyset \qquad \text{[Finite additivity]}$$

Let A^c be function from a credence function and a world to the interval $[0, -\infty]$, and let us call A^c a *credal accuracy function*. Credal accuracy functions measure, as the name might suggest, the accuracy of credence functions; the greater the score, the closer to the ideal, omniscient credal state you are, and therefore the better your credence function.[25] Then we can define the following:

Definition 13 (Expected Credal Accuracy)**.**

$$EA^c_{Cr}(Cr') =_{def} \sum_{w \in W} Cr(\{w\}) A^c(Cr', w)$$

[25] We needn't worry at this point about the precise details of how these functions work. There is good discussion in e.g. Pettigrew (2016a), and we will discuss some of these issues in more detail in chapter three.

This gives us the expected credal accuracy of Cr, relative to Cr'. We get this value by, for all $w \in W$, looking at how likely Cr' thinks it is that we're at w, then looking at the credal accuracy score Cr receives at w, and then summing those weighted values. We can now define propriety.

Definition 14 (Propriety)**.** *A^c is proper iff if Cr is a probability function, then for all Cr', $EA^c_{Cr}(Cr) \geq EA^c_{Cr}(Cr')$.*

Typically, in fact, something even stronger is required in the accuracy epistemological literature as it pertains to epistemic norms for credences: *strict* propriety, according to which, your probabilistic credence function regards itself as *uniquely* maximising expected accuracy.

Definition 15 (Strict Propriety)**.** *A^c is strictly proper iff if Cr is a probability function, then for all Cr', $EA^c_{Cr}(Cr) \geq EA^c_{Cr}(Cr')$ with equality iff $Cr = Cr'$.*

As the names suggest, strict propriety entails propriety. Since propriety and strict propriety will appear a few times in this thesis, it's worth spending some time discussing them. There are a few connected ways to motivate (strict) propriety. These all (as far as I'm aware) exploit the fact that without (strict) propriety, there can be other credence functions that have at least as much, or even greater, expected accuracy than your own credences. It's not difficult to draw some undesirable consequences from this fact, though exactly what the most damning results are is probably not completely settled. One idea is that if another credence function has greater (at least as great) expected accuracy than (as) your own, then there would seem to be nothing wrong with just adopting that alternative credence function, despite not receiving any evidence. But, the idea goes, changing your mind without receiving new evidence is not a rational thing to do. So strict propriety is needed to forbid this kind of irrational change in view (something like this is at

work in Oddie (1997), and Greaves and Wallace (2005, pp. 420-22)).[26] A related thought comes from David Lewis, though it uses an ideology that prefigures anything accuracy-theoretic. He asks us to imagine a magazine *Consumer Reports*, which in addition to ranking toasters, dishwashers and the like, ranks consumer affairs magazines — and moreover, does not rank itself top. It's quite tempting to think that we should not listen to the advice of *Consumer Reports* when shopping for a new kettle, since it undermines its own advice. On the face of it, having an accuracy function which is not (strictly) proper can put our own beliefs in the position of Consumer Reports: giving us a perspective on the world, while simultaneously undermining that very perspective (D. Lewis, 1971). In Lewis's idiom, our 'inductive methods' should be *immodest*. This idea that rational beliefs should be immodest has itself been leveraged in favour of strict propriety, manifesting itself in at least a couple of ways. One idea is that regarding some other beliefs as more accurate than your own puts you in a sort of Moore-paradoxical state (a state where you believe p, but also believe that you don't believe p) — where it's taken for granted that to be in such a state a) amounts to your perspective undermining itself and b) is therefore bad news (Joyce, 2009, p. 277). Another is that, since rational beliefs should always regard themselves as best, and since in principle any probability function could in the right circumstances be rational, we need strict propriety to secure this feature of the accuracy scores (Joyce, 2009, pp. 227-29; Gibbard, 2008).

It would be inaccurate to say the principle has no doubters, or faces no challenges — what philosophical principle doesn't — but it is very widely endorsed.[27]

[26]There are some views that conceivably have this consequence. For instance, certain versions of the permissivist thesis that some bodies of evidence permit multiple, differing rational reactions to it may in turn permit you to change your mind without any change in evidence: precisely because the evidence is permissive. Such cases remain very controversial, and that such evidence-free changes in belief could be permitted is sometimes leveraged as an objection *to* permissivism, as in White (2005).

[27]Indeed, Pettigrew (2011, 2.1.1) points out that it is unclear whether these latter strands of

Kevin Dorst has described strict propriety as a 'standard operating procedure' in the context of the accuracy epistemology literature (Dorst, 2019, p. 181). For the briefest of samples, it's made use of, either directly or indirectly, in Joyce (1998); Greaves and Wallace (2005); Leitgeb and Pettigrew (2010b); Pettigrew (2016a); Schoenfield (2017); Briggs and Pettigrew (2020) — and many, many more places besides. Nor is it some spandrel; that a credal accuracy function is strictly proper is a crucial part of credal accuracy arguments. What's more, as Campbell-Moore and Levinstein (2021) have recently shown, the weaker principle, propriety, entails strict propriety when combined with the following principle.

Definition 16 (Truth-directedness)**.** *If Cr and Cr' are credence functions, and $Cr(p) \geq Cr'(p)$ for all true propositions at w, $Cr(p) \leq Cr'(p)$ for all false propositions at w, and either $Cr(p) > Cr'(p)$ for some true proposition at w or $Cr(p) < Cr'(p)$ for some false proposition at w then $A^c(Cr, w) > A^c(Cr', w)$.*

Truth-directedness is arguably constitutive of accuracy epistemology, since it encodes the idea that credences which are closer to the truth rather than further away are epistemically better. It's not clear one could endorse the veritist value theory underlying accuracy epistemology while rejecting truth-directedness. In which case, even propriety — which I am not sure anyone denies — is enough to get accuracy epistemologists all the way to strict propriety.

There is an appealing analogy between proper credal accuracy functions and $\cap B$-coherence.[28] Proper credal accuracy functions require that credence functions regard themselves as maximising expected accuracy. More informally, the assumption that legitimate credal accuracy functions are proper amounts to the

argument can get strict propriety, or just propriety. Given propriety, you could only ever regard some other perspective as as good as your own, but never better, thereby blocking 'Moorean' or 'modest' consequences like the above.

[28]Thanks to Kenny Easwaran for pointing out that it was propriety rather than strictness doing the work in this argument.

requirement that your credences shouldn't do epistemically badly by their own lights. In the context of maximising expected accuracy, this amounts to not doing badly at those worlds your credence function treats as live. And this is exactly analogous to what $\cap B$-coherence requires: that your beliefs not do epistemically badly by their own lights, by not doing badly at those worlds your belief set treats as live. Both requirements impose a kind of immodesty, in Lewis's terminology, and belief-sensitivity, at a level of coarseness appropriate to the attitude they are concerned with.

Because of this analogy, there is strong *prima facie* pressure to accept $\cap B$-coherence as a genuine requirement in the case of outright belief. For arguments that tell against $\cap B$-coherence will likely have to tell against the idea that your credences should do best by your own lights. In doing so, they will likely tell against propriety. But to do so is to risk the many goods of the credal accuracy programme, which as we mentioned earlier, virtually always assume propriety somewhere.

The structural analogy with propriety strongly suggests that our accuracy argument for deductive cogency is on the right track. It also applies pressure to W-coherence. W-coherence makes us care about every world. But sometimes, we'll have ruled some of those worlds out (as $\cap B$-coherence, and in fact, propriety, allows). So defenders of W-coherence need to explain why we should worry about *those* worlds.[29] As we'll see before long, it's just not clear that they can

[29]Kenny Easwaran provided me with the following interesting observation: (strict) propriety serves as a constraint on scoring rules because it's tacitly assumed that rational credence maximises expected accuracy by its own lights. This leads to probabilism, as do the credal accuracy dominance arguments found in e.g. Pettigrew (2016a), which use the analogue of the strong W-dominance condition defined earlier. The credal accuracy dominance arguments require that you not be dominated with respect to *every* world, making them analogous to the W-coherence arguments we saw above. However maximising expected credal accuracy means ignoring those worlds your credence function has ruled out, making those arguments analogous to $\cap B$-coherence. In the credal case, both lead to probabilism (in fact, since being non-probabilistic means being strongly W-dominated, this also holds for the credal analogue of $\cap B$-dominance arguments). But in the case of outright belief the two come apart, as we've seen — a surprising result.

provide such an explanation.

1.5.2 Pairwise consistency

Before then, there's another problem to consider. This issue for W-coherence relates to whether it can secure pairwise consistency. As mentioned above, pairwise consistency is an exceedingly plausible requirement of epistemic rationality. If your beliefs are not pairwise consistent then your beliefs contain two propositions that are manifestly inconsistent. It would be bad news indeed if W-coherence could not ensure that your belief set does not contain two manifestly inconsistent propositions. Now proposition 6 told us that if $|F| > T > 0$, then if B is not pairwise consistent, B is strongly W-dominated. This guarantees that W-coherence requires pairwise consistency *under the constraints on A^b described above*. However we might well want to ask why that restriction is imposed. Why must it be that the badness of a false belief is greater in magnitude than the goodness of a true belief? This explanatory challenge has been posed by Steinberger (2019). If $T \geq |F|$ then W-coherence does not guarantee pairwise consistency: there can be W-coherent belief sets where $p, q \in B$ and $p \cap q = \emptyset$. For instance, if $W = \{w_1, w_2\}$, $T = 10$ and $F = -1$, then the belief set $\{W, \{w_1\}, \{w_2\}\}$ is neither weakly nor strongly W-dominated.

Despite this, it is far from clear how the proponent of W-coherence should justify the restriction that $|F| > T$. It's natural to conceive of the difference between F and T as encoding different attitudes to risk. If T is greater than $|F|$ then you are epistemically risk-seeking; you prize the goodness of believing truths over the badness of believing falsehoods. Likewise, if $|F|$ is greater than T, you are epistemically risk-averse; you prize the goodness of avoiding false beliefs over the badness of missing out on true ones. But if we do take T and F to represent as described, it looks difficult to justify the restrictions on F and T.

This is because agents may take different attitudes to epistemic risk, just as they may take different attitudes to practical risk. Perhaps attitudes to risk should not be encoded in the utility function at all. But then it is hard to see why even very even-handed accuracy functions like $T = |F|$ should be forbidden.[30]

One sometimes gets the impression that it is simply supposed to be an upshot of wanting W-coherence to secure pairwise consistency that $|F| > T$.[31] That might be acceptable if W-coherence were the only notion of accuracy coherence available for outright belief, and we were trying to determine its properties. Obviously, once we realise there are rival conceptions of accuracy coherence for belief — as we have done here — that is no longer legitimate.

All of this would be moot if $\cap B$-coherence also only secured consistency relative to a constrained set of accuracy functions. But as we saw above, $\cap B$-coherence secures consistency, and therefore pairwise consistency, regardless of what values for T and F one chooses (at least so long as they respect the constraint described earlier that $T > 0 > F$). Consequently, proponents of $\cap B$-coherence have a far easier time than proponents of W-coherence in explaining why pairwise consistency holds as an epistemic norm.

1.5.3 Single-premise closure

The third problem for W-coherence rests on its connection to single-premise closure. As we've seen, if you are to be either $\cap B$-coherent or W-coherent, then

[30] An alternative would be to handle risk at not at the level of the utility function, but at the level of the decision rule, as in Pettigrew (2016b). We discuss some of these issues pertaining to epistemic risk in section 1.7.2.

[31] Dorst (2019, p. 185) is very explicit on this point, though the idea is also at work in Levi (1967). Steinberger (2019, pp. 662-65) was the first to challenge the (lack of good) arguments for the assumption of conservatism, i.e. that $|F| > T$. Interestingly, Roy Sorensen also assumes that $|F| > T$, despite not being someone particularly associated with epistemic utility theory, and in connection with the badness of believing Moorean propositions rather than the goodness of pairwise consistency *per se* (Sorensen, 1988, pp. 37-38). We return to blindspots in the next chapter.

your beliefs must be single-premise closed. This was an important result, because we argued that single-premise closure was a non-negotiable consequence of any plausible kind of accuracy coherence. However there's an important philosophical problem with the notion of W-coherence; a problem that as yet, we've not discussed.

On the face of it, this notion of W-coherence is quite natural, but we've still said very little about how the W-dominance conditions should be understood. What, in other words, is the role of W? Why do we evaluate B at all worlds in W rather than some subset of W? We are given a clue by Easwaran here:

> I will call the objects of belief "propositions", and I will assume that these propositions are characterized by sets of possible "situations". These situations are not metaphysically possible worlds — for instance, there may be situations in which Hesperus is not Phosphorus. Rather, these situations represent the uncertainties for the agent. A proposition that contains all the situations is one that the agent is certain of. (Easwaran, 2016, p. 819)

So our space of propositions is constructed out of the w in W. But by the same token, these w are just those worlds which our uncertainties leave open. It may not be clear what 'uncertainties' means here. Does it mean that for all w, $Cr(\{w\}) > 0$? It does not, for in his (2016), Easwaran is attempting to reduce credence to outright belief. Within that paper, at least, he does not officially countenance the mental state of credence. With this in mind, we should regard him as describing those worlds which are epistemically open; which are left open by your evidence.

And independently of Easwaran exegesis, this motivation seems reasonable. Why should an agent be concerned about the accuracy of beliefs at worlds which aren't epistemically or evidentially open to them? By their own lights those worlds

are irrelevant.[32]

But in fact, there is a problem with the notion of W-coherence as understood. For now W is being required to play two roles — two roles which should be kept distinct. First, W plays a logical role in determining the set of propositions. In this framework, after all, propositions are just sets of worlds. Second, W plays an epistemic role; it tells us which worlds are left open by our evidence; our 'uncertainties'. But it is not always true that the set of all worlds is the set of worlds left open by our evidence. Indeed, if we have any non-trivial evidence whatsoever then that evidence will exclude some worlds.

With this observation in mind, we can see that the W-dominance conditions are a special case of some much more general dominance conditions: conditions which require you to evaluate your belief set at every world left epistemically open to you. We need this generalisation if we are to accommodate the possibility of learning.[33] Let $E \subseteq W$ be the agent's evidence. Then our generalised W-dominance conditions (absent a better name: *E-dominance)* say the following:

Definition 17 (Strong E-dominance)**.** *A belief set B' strongly E-dominates another belief set B iff $A^b(B', w) > A^b(B, w)$ at every $w \in E$, where $E \subseteq W$.*

Definition 18 (Weak E-dominance)**.** *A belief set B' weakly E-dominates another belief set B iff $A^b(B', w) \geq A^b(B, w)$ at every $w \in E$, and moreover, $A^b(B', w) > A^b(B, w)$ for at least some $w \in E$, where $E \subseteq W$.*

These are put to work in the obvious corresponding normative claim:

[32] At least in some sense of 'lights'; exactly what this means will depend on our understanding of evidence, which we'll remain silent on here. We discuss differing conceptions of evidence more in chapters four and five.

[33] Note: we can stay silent about exactly how learning happens. That is, we don't need to provide an outright belief analogue of conditionalisation. However the dynamics turn out to work, we only need the very weak assumption that sometimes, upon learning, you rule out some worlds. We'll have some more to say about belief dynamics in section 1.7.

> **E-coherence:** It's rationally prohibited for B to be either strongly or weakly E-dominated.

E-coherence shares many attractive properties with W-coherence and $\cap B$-coherence. For example, the same reasoning we used to show that they require belief in W can be extended to E-coherence. Belief in W will make you better off at every world, and so will make you better off at every world in E.[34] Moreover, E-coherence has the desirable consequence that it requires belief in E. The reasoning is similar. Belief in E makes you better off at every world in E, so if a belief set lacks E, it is E-dominated by an otherwise identical belief set that also includes E.[35]

E-coherence can secure pairwise consistency, at least subject to some constraints.[36] The more pressing issue is whether E-coherence still secures single-premise closure. And it turns out that it does not:

Proposition 19. *Single-premise closure is not necessary for E-coherence.*

Proof. There is no especially elegant way to prove this, so I provide a simple toy model in which there is an E-coherent belief set that is not closed under single-premise entailment.

Let $W = \{w_1, w_2, w_3\}$ and $E = \{w_1, w_2\}$. When $|W| = 3$, $|\mathcal{P}(W)| = 8$ and $|\mathcal{P}(\mathcal{P}(W))| = 2^8 = 256$. However only 2^7 of those belief sets contain W; only 2^6 both contain W and do not contain $\emptyset$; only 2^5 contain both W and E, and not $\emptyset$; only 2^4 contain both W and E, and neither $\emptyset$ nor $W \setminus E$. These are described in table 1.1.

[34] Parallel reasoning establishes that E-coherence prohibits belief in the contradictory proposition.

[35] Parallel reasoning establishes that E-coherence prohibits belief in the complement of your evidence.

[36] The issue describing the values T and F we saw in section 1.5.2 recurs, for W-coherence is just a special case of E-coherence.

Belief set	$\boldsymbol{A^b(\cdot, w_1)}$	$\boldsymbol{A^b(\cdot, w_2)}$
$\{W, E\}$	$2T$	$2T$
$\boldsymbol{\{W, E, \{w_1\}\}}$	$\boldsymbol{3T}$	$\boldsymbol{2T + F}$
$\{W, E, \{w_2\}\}$	$2T + F$	$3T$
$\boldsymbol{\{W, E, \{w_1, w_3\}\}}$	$\boldsymbol{3T}$	$\boldsymbol{2T + F}$
$\{W, E, \{w_2, w_3\}\}$	$2T + F$	$3T$
$\{W, E, \{w_1\}, \{w_1, w_3\}\}$	$4T$	$2T + 2F$
$\{W, E, \{w_2\}, \{w_2, w_3\}\}$	$2T + 2F$	$4T$
$\{W, E, \{w_1, w_3\}, \{w_2, w_3\}\}$	$3T + F$	$3T + F$
$\{W, E, \{w_1\}, \{w_2\}\}$	$3T + F$	$3T + F$
$\{W, E, \{w_1\}, \{w_2, w_3\}\}$	$3T + F$	$3T + F$
$\{W, E, \{w_1\}, \{w_1, w_3\}, \{w_2, w_3\}\}$	$4T + F$	$3T + 2F$
$\{W, E, \{w_1\}, \{w_2\}, \{w_1, w_3\}\}$	$4T + F$	$3T + 2F$
$\{W, E, \{w_1\}, \{w_2\}, \{w_2, w_3\}\}$	$3T + 2F$	$4T + F$
$\{W, E, \{w_1\}, \{w_2\}, \{w_1, w_3\}, \{w_2, w_3\}\}$	$4T + 2F$	$4T + 2F$
$\{W, E, \{w_2\}, \{w_1, w_3\}\}$	$3T + F$	$3T + F$
$\{W, E, \{w_2\}, \{w_1, w_3\}, \{w_2, w_3\}\}$	$3T + 2F$	$4T + F$

Table 1.1: Belief sets and accuracy scores

The eight belief sets in the second half of the table are not pairwise consistent, and therefore can be ignored — any proponent of E-coherence will want to secure pairwise consistency, and so will only evaluate B with respect to accuracy functions that can secure it. Of the remaining belief sets, it's easy to see that $\{W, E, \{w_1\}\}$ and $\{W, E, \{w_1, w_3\}\}$ score identically. But we can also note that they are E-coherent. The only other belief set which scores as well or better at w_1 is $\{W, E, \{w_1\}, \{w_1, w_3\}\}$. But this belief set scores worse at w_2 and so cannot weakly E-dominate the two just described. Therefore the two sets are both E-coherent. However only one is single-premise closed. It follows that E-coherence does not require single-premise closure. □

It's worth noting why W-coherence gets single-premise closure and E-coherence does not. W-coherence gets single-premise closure because, as we saw in our proof of proposition 7, when B isn't single-premise closed there's another set B'

which does at least as well at every world, and better at some. When B isn't single-premise closed because some p is in B, q is not in B, and $p \subset q$, the belief set B' which swaps p for q but is otherwise identical to B does as well as B at those worlds in p and q, and those in neither, but better at those worlds in q but not p. What happens with E-coherence is that E can exclude the worlds at which B' is better — those worlds in q, but not p. And in cases like this, there's no longer any possible benefit to being single-premise closed.

In our case, notice that for the set which is not single-premise closed, $A^b(\{W, E, \{w_1\}\}, w_3) = T + 2F$. Whereas for the single-premise closed set, $A^b(\{W, E, \{w_1, w_3\}\}, w_3) = 2T + F$. So the latter weakly W-dominates the former; it does equally well at w_1 and w_2 but better at w_3. But because E-coherence only requires that we evaluate those belief sets at w_1 and w_2, the latter belief set is unable to outscore the former at any world. The evaluation ignores the case where the single-premise closed belief set does better. That is why E-coherence cannot require single-premise closure.

So in effect, the situation is as follows: there is a conflation at the heart of W-coherence; a conflation between the logical role of the worlds in W and the epistemic role. We can fix this conflation, and thereby allow for the possibility of learning, by generalising W-coherence to get E-coherence. But, having done so, we can no longer secure single-premise closure. That is a serious cost.

1.5.4 Summary

All three issues in this section show $\cap B$-coherence to be preferable to W-coherence. $\cap B$-coherence is importantly structurally analogous to a central principle in the credal accuracy framework, whereas W-coherence is not; $\cap B$-coherence can secure pairwise consistency under very general conditions, whereas W-coherence cannot; $\cap B$-coherence can secure single-premise closure under very general con-

ditions, whereas W-coherence, and its philosophically-required generalisation E-coherence, cannot. So we should conclude that $\cap B$-coherence is the notion of accuracy coherence to be concerned with; W/E-coherence cannot do the epistemological work required of it.

1.6 Objections and replies

In this section I want to consider some objections to my accuracy argument for deductive cogency, and to my case for endorsing $\cap B$-coherence and rejecting W-coherence as a genuine epistemic requirement.

Objection: The accuracy argument for $\cap B$-coherence is trivial. One response I've received to the arguments in this chapter is that deductive cogency *trivially* falls out of $\cap B$-coherence and the $\cap B$-dominance conditions. The worry is that the conclusion is baked into the premisses. There are a couple of things to say in reply. The first is to acknowledge that there *is* a sense in which consistency and multi-premise closure are built into the dominance conditions. But that is not so surprising; ultimately one cannot get out more than one puts in, and a similar thing holds for W-coherence, pairwise consistency and single-premise closure, in any case.

The more important thing to emphasise is that the reason for endorsing deductive cogency isn't the accuracy argument *per se*. It is instead made appealing by its other features: its structural similarity to the desirable propriety condition; its ability to generate epistemic requirements of the necessary strength (relative to an extremely wide range of accuracy functions, moreover). These are what make $\cap B$-coherence, and by association, deductive cogency, attractive.

For those who remain unsatisfied, I'll now consider one other way to press the triviality worry and show that it doesn't work. We said at the outset that

$\cap B$-coherence was supposed to capture a notion of accuracy coherence *that was sensitive to the agent's beliefs*. It does so by requiring that your beliefs not be dominated with respect to those worlds in their intersection. Now, if there are other accuracy coherence notions that can capture belief-sensitivity without requiring consistency, then the charge that consistency is built into the accuracy argument above might have some force. For the objector could then claim that the belief-sensitive $\cap B$-coherence had been deliberately selected to guarantee consistency (and multi-premise closure). The task, then, for those who wish to push this triviality worry is to find some alternative, belief-sensitive notion of dominance that doesn't require consistency. I'll now briefly discuss a couple of these and suggest they can't work.

One way to go is to swap $\cap B$ for $\cup B$. As we've seen, the former is the set of worlds common to all the propositions you believe. Instead, $\cup B$ is the set of worlds in *any* of the propositions you believe. We could take $\cup B$ to be the set of worlds that a belief set treats as live, and then give alternative dominance conditions, and an alternative characterisation of accuracy coherence.

Definition 20 (Strong $\cup B$-dominance)**.** *B' strongly $\cup B$-dominates B iff $A^b(B', w) > A^b(B, w)$ at all $w \in \cup B$.*

Definition 21 (Weak $\cup B$-dominance)**.** *B' weakly $\cup B$-dominates B iff $A^b(B', w) \geq A^b(B, w)$ at all $w \in \cup B$, and $A^b(B', w) > A^b(B, w)$ at some $w \in \cup B$.*

> **$\cup B$-coherence:** It's rationally prohibited for B to be either strongly or weakly $\cup B$-dominated.

We don't even need to investigate the formal properties of $\cup B$-coherence, because it is philosophically unacceptable.[37] Remember, $\cup B$-coherence was intro-

[37] Although it's fairly easy to see that $\cup B$-coherence almost collapses into W-coherence: if W is not in (non-empty) B, then B will be $\cup B$-dominated, for the same reasons that we saw with e.g. W-coherence. And once W is in B, then B must be undominated with respect to all worlds in W, which is just W-coherence.

duced as an alternative, belief-sensitive accuracy coherence notion. But it does not capture the idea that your belief set should only be evaluated at the worlds your belief set treats as live. Let $W = \{w_1, w_2\}$ and $B = \{W, \{w_1\}\}$. In order to check that B is not $\cup B$-dominated we must then evaluate B at both w_1 and w_2, for $w_2 \in \cup B$. But that is to allow no belief-sensitivity at all, for B rules out w_2 on any reasonable interpretation of 'rules out'.

We've seen that $\cup B$ lets in too many worlds to be plausible, so a more restrictive belief-sensitive notion is needed.[38] First, we introduce a notion of *open possibilities*. We say that w is an open possibility for B iff w is in O_B. And we define O_B as follows:

Definition 22 (Open Possibilities)**.** $O_B = \cup\{p \in B : \forall q \in B, q \not\subset p\}$.

Informally: w is an open possibility for B iff w is in *any* of the strongest beliefs in B; any of the beliefs in B such that no other belief logically entails it. We can then define the natural corresponding dominance conditions.

Definition 23 (Strong O_B-dominance)**.** *B' strongly O_B-dominates B iff $A^b(B', w) > A^b(B, w)$ at all $w \in O_B$.*

Definition 24 (Weak O_B-dominance)**.** *B' weakly O_B-dominates B iff $A^b(B', w) \geq A^b(B, w)$ at all $w \in O_B$, and $A^b(B', w) > A^b(B, w)$ at some $w \in O_B$.*

> **O_B-coherence**: It's rationally prohibited for B to be either strongly or weakly O_B-dominated.

O_B-coherence avoids the problem we saw for $\cup B$-coherence. $\cup B$-coherence required that the belief set $\{W, \{w_1\}\}$ be evaluated not only at w_1, but w_2. But O_B-coherence only requires it to be evaluated at w_1, since $\{w_1\}$ is uniquely the

[38] Thanks very much to Daniel Rothschild and Kenny Easwaran for discussion of this condition.

strongest proposition in the set, and w_1 the only world in it. In this case, then, O_B-coherence treats as live exactly those worlds $\cap B$-coherence treats as live.

But they do not always agree. Assume still that $W = \{w_1, w_2\}$ and that our belief set is $\{W, \{w_1\}, \{w_2\}\}$. According to O_B-coherence, the set of live possibilities is $\{w_1, w_2\}$, for both $\{w_1\}$ and $\{w_2\}$ are such that no stronger proposition entails them. But according to $\cap B$-coherence, the set of live possibilities is the empty set, for there's no world common to every proposition in the set.

O_B-coherence is much more promising than $\cup B$-coherence. It has some important minimal properties, such as requiring you to believe the tautologous proposition.[39] But it has problems. Most importantly, it doesn't require your belief set to be single-premise closed. And as we saw in section 1.3, single-premise closure is a non-negotiable principle of epistemic rationality: any notion of accuracy coherence worth having must be able to secure it.

Proposition 25. *Single-premise closure is not necessary for O_B-coherence.*

Proof. Let $W = \{w_1, w_2, w_3\}$, $T = |F|$, and let $B = \{W, \{w_1, w_2\}, \{w_1\}, \{w_2\}\}$. Therefore $O_B = \{w_1, w_2\}$, for both $\{w_1\}$ and $\{w_2\}$ are the strongest propositions in B. B is not single premise-closed, for although $\{w_1\}$ and $\{w_2\}$ are in B, neither $\{w_1, w_3\}$ nor $\{w_2, w_3\}$ are.

We can show that B is O_B-coherent by looking at the table 1.1. $A^b(B, w_1) = A^b(B, w_2) = 3T + F = 2T$. Any rival belief set that does at least as well at one world does worse at the other. So B is O_B-coherent without being single-premise closed. □

We've seen two variant belief-sensitive notions of accuracy coherence: $\cup B$-coherence and O_B-coherence. Neither work. The former is not really belief-

[39] The argument for this is parallel to the argument that $W/ \cap B/E$-coherence require belief in the tautology. Parallel arguments also establish that O_B-coherence requires belief in O_B, and forbids belief in both the contradictory proposition and $W \setminus O_B$.

sensitive after all, the latter cannot even provide single-premise closure. That they don't work strongly suggests that $\cap B$-coherence does not require consistency *trivially*. Rather, it requires consistency as a result of trying to capture belief-sensitivity in the dominance conditions at all.[40,41]

Objection: Lockeans maximise expected accuracy (Dorst, 2019). And since Lockeans get single premise closure (but not multi-premise closure), there's a principled accuracy-theoretic story we can tell which satisfies your desiderata without leading to deductive cogency. Moreover, maximising expected value is exactly the sort of thing which ideally rational agents should be doing.

Here is how the proposal just mentioned works. Dorst argues in his (2019) that we can get a Lockean thesis out of standard accuracy-theoretic resources. Lockeans say that you believe p iff $Pr(p) \geq t$, where t is some threshold in the unit interval. Accuracy theorists (can) say that you should believe p only if doing so maximises expected accuracy; in other words, if believing p is expectedly as good or better than suspending judgment. Since suspending judgment is guaranteed to have epistemic value of 0, this corresponds to the following. Believe p iff:

[40]Maybe the convinced sceptic could deny that belief-sensitive dominance conditions are worth pursuing. But given the problems with W/E-coherence detailed above, that would be some reason to think the project of deriving norms on outright belief from accuracy dominance reasoning is not worth pursuing. Even in this case, then, we would wind up with an interesting conclusion, for it would turn out that there are non-trivial bounds on the usefulness of accuracy-theoretic reasoning.

[41]There's another sort of triviality worry I'll briefly mention, which is that, just as an inconsistent belief set is strongly $\cap B$-dominated by all belief sets (including itself), it vacuously strongly $\cap B$-dominates all belief sets, too (including itself). Instead of holding that it's *impermissible* to hold a belief set that is dominated by some other belief set, you might instead hold that it is *permissible* to hold any belief set that dominates all others — as any inconsistent belief set does, given the $\cap B$-coherence condition. At least some reason for preferring the former is that that dominance principles are typically interpreted in terms of *prohibitions*, rather than *permissions*, in which case, it's the former understanding — the one I've been working throughout — that we should use. If this isn't persuasive, we could invoke supplementary premisses, though admittedly this would not be especially desirable. Here is an example. Call a belief set 'nihilistic' if it regards no belief sets as better or worse than any other (see Oddie (1997, p. 537) for a similar property). The fairly modest assumption of anti-nihilism would suffice to independently block inconsistent belief sets, even allowing for this alternative way of understanding the dominance conditions. This is because, by the lights of inconsistent belief sets, no belief sets are better than any other. Thanks to Richard Roth for prompting me to think about this.

$$Pr(p)T + Pr(\neg p)F \geq 0$$

Rearranging, we get:

$$Pr(p)T + (1 - Pr(p))F \geq 0$$

And eventually:

$$Pr(p) \geq \frac{|F|}{|F| + T}$$

That is our threshold.

On the face of it, we should expect this Lockeanism to secure single premise closure. After all (and as we saw in section 1.3), if $Pr(p) \geq t$ and $p \subseteq q$, then $Pr(q) \geq t$, by finite additivity. But accuracy-theoretic Lockeanism does not get this result.

Proposition 26. *Single-premise closure is not necessary for maximising expected accuracy.*

A full proof is given in Appendix 1.C, but I'll sketch the counterexample here. Suppose that $T = |F|$ so the prize of being right is exactly proportionate to the cost of being wrong. We saw above that for accuracy-theoretic Lockeans, the threshold is $\frac{|F|}{|F|+T}$, which in this case is 0.5.[42] Suppose that we have three worlds, and let $Pr(\{w_1\}) = Pr(\{w_2\}) = 0.5$. It follows that the expected utility of both $\{w_1\}$ and $\{w_1, w_3\}$ is zero. Since both have zero expected utility, you can maximise utility either by having them in your belief set or not. In particular, you can have $\{w_1\}$ in your belief set, not have $\{w_1, w_3\}$ in your belief set, and still

[42]Similar counterexamples are easily constructed using alternative values, including where $T < |F|$ (and the threshold is thereby greater than 0.5).

maximise expected accuracy. But in doing so, you violate single-premise closure, for $\{w_1\} \subset \{w_1, w_3\}$.

It's crucial for this argument to go through that you give probability zero to w_3: otherwise, it wouldn't be that both $\{w_1\}$ and $\{w_1, w_3\}$ had an expected accuracy of zero. So really, what the argument shows is that if you have any non-trivial information whatsoever, then there are cases in which you maximise expected accuracy while violating single-premise closure. Here is a well-known property of (some) probability functions.

Definition 27 (Regularity)**.** *A probability function Pr is regular iff for all $w \in W$, $Pr(\{w\}) > 0$.*

In other words, then, the argument shows that if your probability function is not regular, we can get the closure violation just mentioned.

Why does accuracy-theoretic Lockeanism get this result? After all, we mentioned earlier on that Lockeans seem to vindicate single-premise closure, not violate it. The reason is that accuracy-theoretic Lockeanism is not really Lockeanism. Lockeans say 'believe p iff $Pr(p) \geq t$'. Accuracy-theoretic Lockeans say 'believe p only if $Pr(p) \geq t$'; they cannot endorse the converse direction. More specifically, while they can endorse the claim that if $Pr(p) > t$, then you should believe, they cannot endorse the claim that if $Pr(p) = t$, then you should believe.[43] For in this case, the score in expectation of believing p is identical to suspending on p. And maximising expected accuracy therefore lets you choose which to do.[44]

[43]Couldn't we tweak the statement of Lockeanism so that it says 'Believe p iff $Pr(p) > t$, and then avoid this problem? We could not, for then maximising expected accuracy would violate the statement in the opposite direction; there would be cases where you believe p (because doing so maximises expected accuracy) even though $Pr(p) = t$.

[44]Nothing deep here hangs on the accuracy theorist's choice of value theory; everything hangs on the fact that there are cases where the expected score of believing is nil, in which case both believing and suspending are permissible. Because of this, similar problems will afflict, for instance, the theory of rational belief in Littlejohn and Dutant (2021), even though they use a 'gnostic' (knowledge-theoretic) value theory, rather than an accuracy-theoretic one. There'll still be cases

Could the accuracy theorist impose regularity on Pr? It's hard to see how, without also ruling out conditionalisation. But conditionalisation is extremely plausible, and supported by accuracy arguments to boot.[45] Dropping it is very unattractive.

It's also worth noting that the proposal to maximise expected accuracy relative to your credence function isn't open to Easwaran. For in his (2016) he is attempting to reduce credence to belief. Therefore beliefs can't come about by maximising expected accuracy with respect to your credences, since for Easwaran, credences are not to be countenanced at that level of explanation. In effect then, Easwaran's attempted reduction will succeed only at the cost of failing to require single-premise closure.

So even maximising expected accuracy is not enough to vindicate single-premise closure.

Objection: Even granting these problems for W-coherence and E-coherence, there are other dominance conditions and coherence notions you've failed to consider. So the accuracy argument for deductive cogency does not go through.

It's true that there are other dominance conditions, and corresponding coherence notions. In fact, there are infinitely many: those which require you not be dominated at w_1; conditions which require you not be dominated at w_1 and w_2, and so on. Some of these dominance conditions may even satisfy the minimal desiderata of securing single-premise closure and pairwise consistency. However it isn't enough that the dominance conditions be formally adequate. They have to be explanatory. And it is hard to see what explanatory rivals we have left out. The

where the expected value of believing is just that of suspending (namely, nil) in which case we can run the same argument as before, thereby showing single-premise closure to not hold in this setting.

[45] Actually, as we'll see in chapters four and five, things are considerably more complicated than this, and it's not clear that accuracy arguments support conditioning rather than some alternative update function. Importantly, though, all the update functions floated as candidates for maximising expected accuracy are compatible with your not having a regular accuracy function post-update.

toy w_1 and w_1-and-w_2-dominance conditions are certainly not informative in this sense, and as we saw above, rival belief-sensitive notions of accuracy coherence like $\cup B$-coherence and O_B-coherence may be explanatory, but have problems of their own. Our $\cap B$-dominance conditions, and $\cap B$-coherence, are adequate in the sense just mentioned, for they generate epistemic requirements of the necessary strength, and moreover they are explanatory — in no small part due to the structural analogy with propriety; $\cap B$-coherence requires doing as well as possible by your own lights. The other explanatory rivals we've considered: W-coherence, E-coherence, O_B-coherence, have all turned out to be inadequate.

Objection: There are $\cap B$-coherent belief sets that are weakly W-dominated. This is a problem for the claim that $\cap B$-coherence captures an accuracy-theoretic notion of belief coherence.

The first sentence of this objection is true. There are, to anticipate a result from the next chapter (proposition 115), some $\cap B$-coherent sets that are weakly W-dominated. It is not clear to me that the second sentence is true. In the cases of interest, B' weakly W-dominates $\cap B$-coherent B by scoring better than B at some world B rules out. That is, some $w \notin \cap B$. But as far as B is concerned, that world is irrelevant; ignoring that world was part of the point of using a belief-sensitive kind of accuracy coherence. So it does not matter that this alternative belief set scores better at this world; in the same way that atheists should not be concerned that they will suffer in the (by their lights, ruled out) possibility that unbelievers go to hell.

One might press the complaint by appeal to an analogy with certain Dutch Book arguments. For instance, the Dutch Book for regularity appeals to a similar principle: that if your credence function is not regular, you risk losing money; if you have regular credences, you do not (Skyrms, 1980, p. 74). Here is an example. Suppose that you rule out (give credence zero to) w. Then you are permitted to

accept a bet that pays nothing if $W \setminus \{w\}$ is true, and which will require you to be tortured if $\{w\}$ is true — the expected utility of accepting and rejecting the bet are the same, namely, nil. In reply, I think I am pretty happy to bite the bullet here. It is fine to ignore worlds you have ruled out, and consequently, it is fine to take on the risk of torture in the case described above. By your own lights, it is no risk at all. Similarly, it is perfectly okay that there are some $\cap B$-coherent belief sets that are weakly W-dominated, insofar as the worlds at which the alternative belief set scores better than your belief set is one you have ruled out. Once again, by your own lights it is no risk at all.

There is a general issue here, reminiscent of some of the literature on epistemic contextualism and pragmatic encroachment. Take the (canonical) pair of cases from DeRose (1992, p. 591):

> Bank Case A. My wife and I are driving home on a Friday afternoon. We plan to stop at the bank on the way home to deposit our paychecks. But as we drive past the bank, we notice that the lines inside are very long, as they often are on Friday afternoons. Although we generally like to deposit our paychecks as soon as possible, it is not especially important in this case that they be deposited right away, so I suggest that we drive straight home and deposit our paychecks on Saturday morning. My wife says, "Maybe the bank won't be open tomorrow. Lots of banks are closed on Saturdays." I reply, "No, I know it'll be open. I was just there two weeks ago on Saturday. It's open until noon."
>
> Bank Case B. My wife and I drive past the bank on a Friday afternoon, as in Case A, and notice the long lines. I again suggest that we deposit our paychecks on Saturday morning, explaining that I was at the bank on Saturday morning only two weeks ago and dis covered that it was

> open until noon. But in this case, we have just written a very large and very important check. If our paychecks are not deposited into our checking account before Monday morning, the important check we wrote will bounce, leaving us in a very bad situation. And, of course, the bank is not open on Sunday. My wife reminds me of these facts. She then says, "Banks do change their hours. Do you know the bank will be open tomorrow?" Remaining as confident as I was before that the bank will be open then, still, I reply, "Well, no. I'd better go in and make sure."

Many people have the judgment that in the second bank case, we wouldn't know that the bank was open. One diagnosis is that when the stakes are high, previously ruled out possibilities can apparently become 'live', leading you to 'lose' knowledge you otherwise would have possessed (Stanley, 2005). I think something similar is (or may be) driving judgments in the kinds of cases we are interested in, too. Although I don't want to get too into the weeds in the debate over whether there is pragmatic encroachment on knowledge, my hunch is that the answer I prefer in these sorts of cases — that knowledge isn't being lost, and a version of epistemic purism is true — has an analogue in the case of the Dutch Book for regularity, and moreover, for $\cap B$-coherence, too.

1.7 Further discussion

1.7.1 The relationship between belief and credence

There are some other points of interest I'll now discuss. One concerns the relationship between belief and credence. I've mostly avoided taking a view on this so far; I've referred to other people's views, and at points in this paper I have even

appealed to particular probability functions. But nothing has really depended on any particular interpretation for those probability functions; they have just been convenient mathematical objects.

With that said, we can make some suggestive, if not irresistible, remarks about what the picture of belief we have discussed might tell us about rational credence. It is at least somewhat tempting to think it pushes us towards a belief-is-credence-one view, of the kind discussed earlier. The reason is that, as we will see in the next chapter, if B is $\cap B$-coherent then B maximises expected accuracy with respect to the probability function Pr such that $Pr(\{w\}) = \frac{1}{|\cap B|}$ for all $w \in \cap B$ (proposition 63).

This is interesting for a couple of reasons. First, the fact that all $\cap B$-coherent belief sets maximise expected accuracy with respect to some probability function is a highly non-trivial property. It is not shared, for instance, by W-coherence: as Easwaran (2016) points out, there are W-coherent belief sets such that there is no probability function they maximise expected accuracy with respect to.[46] Similarly, it will not be shared by E-coherence, since E-coherence does not guarantee single-premise closure. For a belief set to maximise expected accuracy with respect to a probability function, the 'belief monotonicity' property from fn. 17, wherein if $p \in B$, $Pr(q) \geq Pr(p)$ then $q \in B$, must hold. The failures of single-premise closure for E-coherence then correspond in turn to failures of belief monotonicity, and the possibility of — in general — E-coherent belief sets maximising accuracy with respect to some probability function.

The fact that there is always some probability function that a given belief set maximises expected accuracy with respect to is somewhat appealing if we like the idea that we should not want the perspectives encoded by our outright beliefs, and our credences, to stand in tension. Suppose you have credences. Then if your

[46]The actual examples aren't terribly instructive, so I won't describe them here, but I refer the reader to Easwaran (2016, pp. 846-48) for some discussion.

beliefs don't maximise expected accuracy with respect to them — as in the case of some of the W-coherent belief sets Easwaran has discussed — then your two doxastic outlooks will stand in a sort of tension. From the perspective of your credences, you should have different beliefs. But from the perspective of your outright beliefs, those outright beliefs are (by the analogy with propriety pursued earlier) at least as good as any rival belief set.

There is an even stronger 'integration' of the two doxastic perspectives available, that we may or may not think is desirable. Specifically, we may not want our outright beliefs to treat some world as live that our credences do not, and vice versa. The fact that every $\cap B$-coherent belief set maximises expected accuracy with respect to the probability function assigning the strongest member of B probability one guarantees this. Equally, we may think the relationship between the two outlooks just gestured to is itself too close.

If we want to separate the two perspectives out, it's worth checking whether $\cap B$-coherent belief sets can maximise expected accuracy with respect to probability functions that *don't* assign $\cap B$ probability one. And in fact, there are $\cap B$-coherent sets B and probability functions Pr such that, while the probability function does not give $\cap B$ probability one, B nonetheless maximises expected accuracy with respect to Pr. For example, let $W = \{w_1, w_2\}$, $B = \{W, \{w_1\}\}$, $T = 1$ and $F = -9$. Then $\frac{|F|}{|F|+T} = \frac{9}{10}$, so if $1 > Pr(\{w_1\}) > \frac{9}{10}$, B will maximise accuracy with respect to Pr, even though $\cap B < 1$.

We might conclude that we are being given another way of bolstering Hannes Leitgeb's stability theory of belief. According to the stability theory, you rationally believe p iff your credence in p is (rationally) above some threshold t, and moreover, for all evidence you might learn, your credence in p will remain above some (possibly distinct) threshold r.[47] The idea is that rational belief is *stable*,

[47]This is simplifying slightly; the identity of the set of propositions you might condition upon is something Leitgeb discusses at length (Leitgeb, 2017, ch. 2).

a fact encoded by the requirement that, whatever you learn, your credence in the believed proposition cannot change *too much*. An interesting feature of the stability view is that it combines aspects of both the Lockean view — you believe a proposition only if your credence in that proposition is above some (possibly-non-maximal) threshold — with the requirements of consistency and multi-premise closure. As we saw earlier, puzzle cases like the lottery and the preface often put pressure on this combination of views. Moreover, the stability view holds not only in the trivial case where the Lockean threshold is 1, but in cases where that threshold is sub-maximal (Leitgeb, 2017, p. 87). We might then think we could argue for the stability view as follows: we want our beliefs and our credences to be accuracy-conducive, and moreover, we want their perspectives to align. In particular, we don't want our credences to recommend some alternative set of beliefs. In which case, our credences and beliefs must hang together in the way described above, where the resulting belief set is both consistent and multi-premise closed — as the stability view requires. They key premise is just that we do not want our two doxastic systems to undermine one another.

While this is appealing, I am not sure what to make of the argument just given, because of the following, second fact: there are $\cap B$-coherent sets which *only* maximise expected accuracy with respect to the probability function Pr such that $Pr(\{w\}) = \frac{1}{|\cap B|}$ for all $w \in \cap B$ (proposition 123 in the following chapter). In other words, that existentially quantified claim I made first will no longer be true if we universally quantify it.

Now Leitgeb is opposed to the belief-in-credence-one view. He thinks it is too demanding, and that there are beliefs we hold despite not being certain of them. He appeals to well known counter-intuitive decision-theoretic consequences of the view, some of which are related to things we have already discussed (Leitgeb, 2017, pp. 42-44). For instance, given belief-as-credence-one, we are permitted to

accept a bet in a proposition we believe even if there is no benefit to being correct, and the cost of being wrong is very great. As mentioned before, many people are (or would be) reticent to accept such bets, and one purported consequence of the argument is that we are not *absolutely* certain even of very obvious truths.

The upshot is that Leitgeb wants there to be sub-maximal Lockean thresholds for belief (with the added constraint that those beliefs are 'stable', though we set this aside here). But as we have seen, given the facts described above, such thresholds are not in general guaranteed on our framework. For some propositions, your belief in them will maximise expected accuracy iff you are certain of them. There are points where Leitgeb seems to acknowledge that the sub-maximal thresholds need not be universally available, and that in some contexts it might be okay that you believe iff you have credence one (see Leitgeb, 2017, p. 84). But by itself, this won't solve the problem we are thinking about. Nothing in the examples we have described says anything about the content of the propositions we are thinking of; we have just specified some mathematical models with a particular property. Accordingly, there is no way of guaranteeing that the sub-maximal-thresholds will line up in the right way with particular contexts, as Leitgeb seems to require by allowing that in some contexts the threshold for belief could be credence one. In short, despite some initial appearances, I'm not so sure that our framework does lend any new support to the stability view.

Of course, insofar as people are suspicious of the belief-as-credence-one view, that may be bad news for the proposal I have explored here. But I am not especially persuaded by the standard decision-theoretic arguments levelled against that proposal (for reasons I have already outlined). And in any case, it is not clear rival views of the relationship have a much easier time of explaining accuracy constraints on outright belief, and the relationship between those outright beliefs and your credences. We have just discussed the stability view, and moreover,

the accuracy-theoretic Lockean view does not guarantee your beliefs are single-premise closed. Perhaps we should think there is no systematic connection between belief and credence full stop, as Buchak (2014) has argued. But then there remains the challenge of explaining how your credences can recommend alternative belief sets to your outright beliefs, and the bifurcated perspective represented therein.

1.7.2 Alternative accuracy functions

Another point of interest concerns the accuracy functions we've worked with. These simple functions have been standard in the recent literature; from Easwaran (2016) thereafter. Functions that are at least somewhat similar were appealed to in the initial work on this topic by the likes of Levi and Hempel, as we mentioned earlier (we will mention some differences shortly). Naturally, though, other functions are conceivable. Are other functions preferable? Here, I want to discuss a few possible shortcomings of the way we have set things up, and suggest that our framework does manage to accommodate them. I don't mean this to be the last word on whether the accuracy of outright beliefs should be measured differently. In fact, I think this question remains underexplored (a defect I've not remedied here). That said, I do think that our position is at least defensible.

One worry that Easwaran discusses, and that I've encountered in discussion or presentation of this work, is that the accuracy functions we use are insensitive to differences between the epistemic value of different propositions. For example, let p and q be truths at w. Specifically, let p be the proposition that there are 43 teabags in the jar, and q be some extremely important and exciting claim about fundamental physics. Both are true at w. Yet $A^b(\{p\}, w) = A^b(\{q\}, w)$. Surely a belief in q is more epistemically valuable than a belief in p? In reply, one might have the function accommodate these apparent differences by having

the accuracy function be *proposition-sensitive*, so the epistemic goodness (badness) of believing various propositions when true (false) can vary. The idea would be that instead of some proposition-invariant scores T and F, there are for all p proposition-relative scores T_p and F_p.

Easwaran has shown that in the case of W-coherence, this will cause various problems. Probably most interesting for our purposes is that single-premise closure can fail under fairly modest 'tweaks' to the scores: if $p \subset q$, but not $T_p \geq T_q$ and $F_q \geq F_p$, then it won't hold in generality (Easwaran, 2016, p. 851). Interestingly, no such problems will afflict $\cap B$-coherence. It remains the case that if your belief set is inconsistent, it's vacuously $\cap B$-incoherent; if it's consistent, then every proposition which is a superset of $\cap B$ has a positive value at the worlds in $\cap B$, and so your belief set will be $\cap B$-dominated if it does not contain it. Nothing hangs on the precise positive values those propositions take.

With that said, I'm not completely convinced that the way to deal with these problems is at the level of the accuracy functions. If pushed, I'd probably endorse the view that there's some sense in which facts about fundamental physics are more valuable than facts about days of the week while denying that these should be represented in the accuracy function. More specifically, I'm tempted to think we should deny that the difference is epistemic, and moreover that the epistemic gods do not distinguish between the goodness of believing truths about fundamental physics and truths about days of the week. From a purely epistemic point of view, it is as useful to spend your life gathering truths about e.g. the number of blades of grass in a field as it is truths about the ultimate nature of the universe (at least insofar you gather an equal number of truths about each). Insofar as there is a difference, it is a practical difference — a difference in what we are interested in. I envision inquiry as a partition-setting activity, where cells of the partition correspond to possible answers to the question we are interested in (an

idea that can be found, appropriately enough, in Levi, 1967). Once that underlying partition is set, we can compare belief sets defined over that partition, using the accuracy functions we have worked with throughout. My tentative claim is just that we are typically more interested in (for instance) facts about fundamental physics than trivialities about the amount of grass in a field. Accordingly, these are the questions we evaluate, and the partition our belief sets are evaluated with respect to. But that aside, there is no epistemic injunction to be more interested in facts about fundamental physics than there is in facts about blades of grass. That said, since we can apparently accommodate some proposition weighting given $\cap B$-coherence, I don't and won't defend this picture to the death.

A second and related issue concerns *informativeness*. Suppose we are interested solely in propositions about fundamental physics. Then it is tempting to think that some very fundamental truth is epistemically preferable to some less fundamental truth which follows from the more fundamental one. Yet as we saw earlier, the accuracy function evaluates the goodness of the two propositions the same. Moreover, by the setup, the underlying partition is fixed. How can we make sense of the fact the former belief is in some way better than the latter?

Some theorists have introduced a separate informativeness measure to capture the difference. Levi (1967, pp. 62-64) uses this strategy, as do Dorst and Mandelkern in much more recent work which, although it concerns guessing rather than belief, has a number of affinities with Levi's project (as they acknowledge; Dorst & Mandelkern, in press, section 3.1).

In fact though, the functions we've used, in combination with the requirement of $\cap B$-coherence, do allow us to tell a story about why the more fundamental claim is epistemically better. Suppose p, q are both true at w but $p \subset q$. Since $\cap B$-coherence requires multi-premise closure, then the agent whose strongest belief is p will be epistemically better off at w than the agent whose strongest belief is q,

because $|\{X : p \subseteq X\} > |\{X : q \subseteq X\}|$. This holds only because your belief set is more informative than uninformative; ruling out more rather than fewer worlds leads to greater (objective) epistemic benefit. This should be doubly appealing to 'accuracy first' epistemologists because it means we can reduce other epistemic virtues like informativeness down to what is by their lights the sole epistemic good: accuracy. What's more, since we can accommodate a sort of propositional-weighting, we can even make sense of the idea that might be thought to underpin this worry about informativeness: that the more fundamental proposition (e.g. p) is more epistemically valuable than the less fundamental proposition (e.g. q) that it entails. We could encode this by having $T_p > T_q$.

The final issue I'd like to discuss is due to work by Matthias Skipper, and what he calls the problem of 'belief gambles' (Skipper, 2021). In brief, the accuracy functions we have used, in combination with various natural decision rules, seem to license 'gambling' on beliefs even though you are not very sure of their truth. For instance, for all W-coherence says, there are some beliefs you can hold which are arbitrarily unlikely on your evidence. For example (this follows from proposition 117, next chapter) the belief set that includes the belief you are in a particular world, and all consequences of that proposition, is always W-coherent. This will hold no matter how (un)likely you think it is that you are in that particular world. And although Skipper does not discuss this, the same problem holds for $\cap B$-coherence, too, as we'll see in the next chapter. Both versions of accuracy coherence will always license gambling — viz. believing — that you are in a particular world.

On the face of it, this is pretty troubling. Partly this is because it seems to preclude an appropriate sort of evidence-responsiveness. We don't always have evidence we're in a particular world. But it's always W/$\cap B$-coherent to believe we're in a particular world. So it's sometimes compatible with W or $\cap B$-coherence that

we flout our evidence. And partly because it seems as though a kind of epistemic risk-taking is always permissible given these notions of accuracy coherence, no matter how conservative the utility function itself is. One could even argue that such 'belief gambles' are the only legitimate options for agents, since they are the only belief sets which will be undominated given any values for T and F whatsoever (see propositions 75 and 117, next chapter). For instance, if an agent is unsure which values T and F take there is a sense in which they cannot go wrong by 'gambling' in the way just described. (We discuss this kind of uncertainty further in chapter three).

I think this is a pretty serious problem. Though I don't have anything fantastically compelling to say, I can gesture at least a couple of replies. First, the coward's answer: this issue afflicts both W-coherence and $\cap B$-coherence. So, insofar as we are interested in adjudicating between those two kinds of accuracy coherence, the matter is not probative. This is helpful to an extent, though one might instead draw the conclusion that this accuracy-theoretic approach to the epistemology of outright belief should be scrapped. I would urge at least a little caution here. The kinds of accuracy functions and dominance reasoning we used are reasonably natural, and if we were going to scrap the project wholesale it would be desirable to have at least some story of where those assumptions are going (apparently, quite badly) wrong.[48]

The second reply is to constrain the accuracy functions. If we ensure $|F|$ is sufficiently greater than T, then (for instance) the expected utility maximising threshold $\frac{|F|}{|F|+T}$ gets greater, and one cannot maximise expected accuracy by believing things of very low probability. As mentioned earlier, just setting $|F| > T$ is enough to ensure that — as in Skipper's case of interest — one cannot maximise expected accuracy by believing something that has probability of just 0.5. The dif-

[48] Skipper (2021, section 2.1) does look at some alternative, risk-sensitive decision rules; I agree that this doesn't help matters.

ficulty with this response, as Skipper points out (following Steinberger, 2019) is that it is not clear what grounds these constraints on the accuracy function.

I think the best answer we can give is that nothing *forces* a choice of T and F such that $|F| > T$. Instead, we select values that correspond to your attitude to risk. The greater T is with respect to $|F|$, the greater your inclination to take risks, and therefore the more you are inclined to 'gamble'. The intuitive oddness of gambling is to be explained by the agents involved being risk-seeking. This response has the disadvantage that even for the very natural accuracy function, on which $T = |F|$, you will often be permitted to gamble. In reply, I would point out that natural accuracy functions are also known to behave strangely in the case of credal accuracy functions. If you measure the accuracy of your credences with the absolute value measure, there are non-probabilistic credence functions which are not accuracy-dominated (Maher, 2002; Pettigrew, 2016a, pp. 34-35). Naturalness is an important desideratum but not the only one.

This response has the advantage that some in the accuracy literature have already been at least somewhat receptive to epistemic risk-seeking (or risk-avoiding).[49] Moreover, it's a familiar point that the resulting kind of evidential permissivism can fall out of allowing differing attitudes to epistemic risk (Kelly, 2013).

So far, everything I've said in this second reply has been equally available to the proponents of both W-coherence and $\cap B$-coherence. However one advantage $\cap B$-coherence has is that it will at least forbid gambles that contain contradictory beliefs. If you gamble and pick the beliefs that maximise expected accuracy, but those beliefs are inconsistent, you'll violate the condition against having bifurcated doxastic perspectives we introduced when discussing Leitgeb's stability view earlier. The W-coherent risk-seeker, on the other hand, may turn out to be-

[49] See e.g. Pettigrew (in press-b, p. 2); Levinstein (2017b, p. 346); my attention was drawn to the latter by Horowitz (2017, p. 272). See Campbell-Moore and Salow (2020b, 2020a) for discussion of epistemic risk in connection with updating.

lieve both a proposition and its negation. So this line of response, tentative though it is, does weakly favour $\cap B$-coherence.

1.7.3 An analogy with inconsistent preferences

Another point of interest arising from the foregoing connects with Dutch Book arguments. When agents are Dutch-bookable their credences commit them to a collection of bets which are guaranteed to make them worse off; moreover, there are alternative credences which would not commit them to this set of bets. The irrationality of such agents is sometimes explained as their having inconsistent preferences: preferring, in light of their credences, to make each of the bets individually, while also preferring not to make the set of bets as a whole. Pettigrew (2020, section 6.2) suggests it is not clear why having inconsistent preferences is irrational. He considers and rejects two reasons for thinking so, one of which is an analogy with belief. He notes that some philosophers have claimed that belief aims at truth (e.g. Wedgwood, 2002) and accordingly that it is wrong to have inconsistent beliefs because in doing so, you guarantee that not all of your beliefs can satisfy that aim. You can suggest by analogy that preferences aim at being satisfied (Pettigrew does not put things exactly this way), and therefore that what is wrong with inconsistent preferences is that they cannot all be satisfied at once. Pettigrew then suggests that what Fitelson and Easwaran (2015) show is that no such thing is true for belief, and accordingly, no such thing can hold for preference, either.

As we've seen, that conclusion is too quick. There are good grounds for thinking that, in fact, it should be possible for all of your beliefs to be true at once. If we respect the analogy, then we should hold that it should be possible for all of your preferences to be satisfied at once. Of course, defending such an analogy will not be straightforward, and I don't mean to undertake doing so here. For one

thing, it's not so clear that your preferences represent a 'perspective', and therefore a perspective that ought to regard itself as best by its own lights (as in the case of belief). But we can at the very least show that this argument against the irrationality of inconsistent preferences is a little too hasty.

1.7.4 Updating

The last thing I'll mention concerns belief dynamics rather than belief statics. Our discussion so far has concerned synchronic constraints on (outright) belief. But there are also a host of interesting philosophical questions about diachronic constraints regarding how one should update when one receives new information. The main theory of updating for outright belief in the literature is known as the 'AGM' model, named for its progenitors, Carlos Alchourrón, Peter Gärdenfors, and David Makinson. There are many things that could and should be said about the relationship between the theory we have outlined here, and the AGM theory of revision. I will say some of them in chapter six, where I discuss topics for future research. For now, I will make just one point.

Informally, the heart of the AGM model is that, upon learning a proposition E, your prior belief set B should be revised to B' where B' is deductively cogent, includes E, and changes B as little as possible (this notion of 'minimal change' is made precise in various ways, which we set aside here; Shear & Fitelson, 2019). Formally, AGM updates are encoded in a belief revision operator $*$, a two place function taking a belief set B and a proposition received as evidence E, and returning another, possibly distinct belief set B', such that $B*E$ is the revision of B upon receiving new evidence E. Various constraints are then placed on $*$, in order to narrow down the space of possible functions to those with adequate properties. There are several 'core' postulates standardly imposed on the operator. For our purposes, it's notable that two of those constraints amount to the requirement that

(posterior) belief sets — i.e. the belief sets one ends up with post-update — must be deductively cogent (Shear & Fitelson, 2019, pp. 499-500). In which case, $\cap B$-coherence, which requires deductive cogency *generally*, is on the face of it going to be an easier fit with AGM updates than the other forms of coherence we've discussed — which do not.[50] If all post-update states must be deductively cogent, it is hard to see how W-coherence, which permits failures of deductive cogency, can be a genuine constraint on belief sets *at a given time*. As we've seen above, W-coherence only requires consistency under certain constraints on A^b, and can never *require* multi-premise closure. But given AGM, all post-update states must be multi-premise closed — something difficult to explain if the form of accuracy coherence one must appeal to is W-coherence.

1.8 Conclusion

This paper argues that considerations of accuracy lead to strong constraints on the set of your beliefs: both consistency and multi-premise closure, or in other words, deductive cogency. It argues that a version of accuracy coherence that can generate those requirements is needed, because alternative notions of accuracy coherence, properly understood, cannot even generate very minimal constraints like single-premise closure. The choice, then, isn't between multi-premise closure (and consistency) on the one hand, and single-premise closure (and pairwise consistency) on the other; it's between multi-premise closure, and no closure constraints at all. Taking this account of belief coherence seriously raises other interesting questions, including the relationship between belief and credence, and how we measure the accuracy of outright belief.

[50] In fact, Shear and Fitelson (2019, pp. 497-98) also argue that AGM requires deductive cogency as a general *synchronic* constraint, too — on prior belief sets, in addition to posterior belief sets.

Appendix 1.A Ultra-conservative W-coherence guarantees consistency

We assumed at the start that, for all A^b, $T > 0 > F$. We now describe two particular classes of accuracy function.

Definition 28. *A^b is conservative iff $|F| > T$*

Definition 29. *A^b is ultra-conservative iff $|F| > (|W| - 1)T$.*

In this section we prove that when A^b is ultra-conservative, W-coherence requires consistency. To prove this, it's helpful to introduce a new notion.

Definition 30 (Aggregate Accuracy)**.** *The Aggregate Accuracy AA^b of B is a two place function $\mathcal{P}(\mathcal{P}(W)) \times \mathcal{P}(W) \to \mathbb{R}$ mapping B to the sum of its accuracy scores at every $w \in E$, where $E \subseteq W$.*

$$AA^b(B, E) = \sum_{w \in E} A^b(B, w)$$

In this paper, we set aside any normative implications of aggregate accuracy *per se*. In what follows we appeal to it as a purely technical construction.

Lemma 31. *If B' either weakly or strongly W-dominates B, then $AA^b(B', W) > AA^b(B, W)$.*

Proof. For B' to weakly W-dominate B, it must be that $A^b(B', w) \geq A^b(B, w)$ at all $w \in W$, and $A^b(B', w) > A^b(B, w)$ for at least some $w \in W$. Since $AA^b(B', W)$ is just the sum of $A^b(B', w)$ for all $w \in W$, if $A^b(B', w) \geq A^b(B, w)$ at all $w \in W$ and $A^b(B', w) > A^b(B, w)$ at some $w \in W$, it follows that $AA^b(B', W) > AA^b(B, W)$. □

We next require the result that:

Lemma 32. *If A^b is ultra-conservative, B is consistent and $\cap B \not\subseteq p$, then $AA^b(\{p\}, \cap B) < 0$.*

In other words, when p is not true at all $w \in \cap B$, then if A^b is ultra-conservative, $\{p\}$ receives negative aggregate accuracy relative to $\cap B$.

Proof. To see this, first note that since $\cap B \not\subseteq p$, there is some $w \in \cap B$ such that $w \notin p$. At this w, $A^b(\{p\}, w) = F$.

But since A^b is ultra-conservative, $|F| > (|W| - 1)T$, and $0 > F$. Since for all B, $\cap B \subseteq W$, and $\cap B \not\subseteq p$, $p \neq W$ and p has at most $|W| - 1$ members. It therefore receives T in at most $|W| - 1$ worlds. But even in this case, $(|W| - 1)T + F < 0$, and therefore $AA^b(\{p\}, \cap B) < 0$. □

Lemma 33. *When A^b is ultra-conservative, B is consistent and $\cap B \not\subseteq p$ then $AA^b(B \cup \{p\}, \cap B) < AA^b(B, \cap B)$.*

Proof. Let A^b be ultra-conservative, B be consistent, $B' = B \cup \{p\}$ and p be such that $\cap B \not\subseteq p$. Equivalently, p is false at some $w \in \cap B$.

By lemma 32, $AA^b(\{p\}, \cap B) < 0$. By additivity of aggregate accuracy, $AA^b(B \cup \{p\}, \cap B) = AA^b(B, \cap B) + AA^b(\{p\}, \cap B)$. Therefore since $AA^b(\{p\}, \cap B) < 0$, it follows that $AA^b(B, \cap B) > AA^b(B \cup \{p\}, \cap B)$. □

Lemma 34. *When A^b is ultra-conservative, the belief set $\{W\}$ is neither weakly nor strongly W-dominated.*

Proof. We prove this by showing that, when A^b is ultra-conservative and $\{W\} \neq B$, there is no B such that $AA^b(B, W) > AA^b(\{W\}, W)$. Therefore by lemma 31 $\{W\}$ is neither weakly nor strongly W-dominated.

First, suppose that $B = \{W\} \cup B'$, where B' is non-empty. In other words, B is given by adding some propositions to $\{W\}$. However then by lemma 32 it follows that $AA^b(\{W\}, W) > AA^b(B, W)$, since when A^b is ultra-conservative, the

addition of any non-tautologous proposition reduces aggregate accuracy. Therefore no superset of $\{W\}$ can either weakly or strongly W-dominate $\{W\}$.

If B is not a superset of $\{W\}$ then $W \notin B$. But then B is strongly W-dominated by $B \cup \{W\}$, which we have already seen cannot either weakly or strongly W-dominate $\{W\}$. Therefore when A^b is ultra-conservative, $\{W\}$ is neither weakly or strongly W-dominated.

□

Lemma 35. *When* $B = \{p : |p| = |W| - 1\}$, $\cap B = \emptyset$

Proof. Every proposition with $|W| - 1$ members is the complement of some singleton $\{w\}$. Therefore at every $w \in W$, there is some $p \in B$ false at w — namely, the complement of $\{w\}$. Since there is no w such that, for all $p \in B$, $w \in p$ it follows that $\cap B = \emptyset$. □

Lemma 36. *When* A^b *is ultra-conservative,* $\{W\}$ *strongly* W*-dominates the belief set* B'*, where* $B' = B \cup \{W\}$*, and* $B = \{p : |p| = |W| - 1\}$.

Proof. At each $w \in W$, $|W| - 1$ members of B are true and one is false (namely, the complement of $\{w\}$). Therefore:

$$A^b(B, w) = (|W| - 1)T + F$$

Since A^b is ultra-conservative, $|F| > (|W| - 1)T$, and therefore $(|W| - 1)T + F < 0$. This means that $A^b(B, w) < 0$.

However $A^b(\{W\}, w) = T$. Since $A^b(B, w) < 0$ at all $w \in W$, and $A^b(B', w) = A^b(B, w) + A^b(\{W\}, w)$, it follows that $A^b(\{W\}, w) > A^b(B', w)$ at all $w \in W$. In other words, $\{W\}$ strongly W-dominates B'. □

Lemma 37. *If* B *does not contain all of the* $(|W| - 1)$*-membered propositions, and* B *is single-premise closed, then* B *is consistent.*

Proof. Suppose that for some w, $W \setminus \{w\} \notin B$. Then for all $p \in B$, $w \in p$. For if $w \notin p$, then $p \subseteq W \setminus \{w\}$. Since B is single-premise closed it would follow that $W \setminus \{w\} \in B$, contrary to our supposition. Since $w \in p$ for all $p \in B$, it follows that $w \in \cap B$ and therefore B is consistent. □

We are finally in a position to show that the W-dominance conditions, combined with an ultra-conservative accuracy function, lead to consistency.

Proposition 38. *When A^b is ultra-conservative, then if B is inconsistent then B is either weakly or strongly W-dominated.*

Proof. Let B be inconsistent. By the contrapositive of lemma 37 B therefore either fails to be single-premise closed, or B includes all $|W|$ of the $(|W| - 1)$-membered propositions.

By proposition 7, if B is not single-premise closed then it is weakly W-dominated. By lemma 36, if B contains all of the $(|W| - 1)$-membered propositions, then it is strongly W-dominated. So either way, inconsistent B is either weakly or strongly W-dominated.[51] □

In effect, what this shows is that when you attach a significantly large disvalue to falsehood, the W-dominance conditions require your belief set to be consistent. So we can get halfway to deductive cogency, just by appealing to a certain kind of very conservative accuracy function.

[51] This is closely related to Fitelson and Easwaran's (2015) Theorem 3. However they assume that B is strongly opinionated, i.e., for all p either p or $\neg p \in B$. We do not assume that.

Appendix 1.B W-coherence never requires multi-premise closure

We might wonder whether attaching greater disvalue to falsehood also secures multi-premise closure. Were that so, then W-coherence would require deductive cogency at least relative to those accuracy functions which greatly disvalue falsehood. But in fact, no degree of epistemic conservatism — indeed, no accuracy functions whatsoever — combine with W-coherence to secure deductive cogency. To show this we need the following result due to Easwaran:

Theorem 39. *If Pr is regular and for all $p \in B$, $Pr(p) \geq \frac{|F|}{|F|+T}$, then B is neither weakly nor strongly W-dominated.*[52]

We can use this fact to show that:

Proposition 40. *Multi-premise closure is not necessary for W-coherence.*[53]

Proof. Let $|W| > 2$ and Pr be a regular probability function such that for some $w \in W$, $Pr(\{w\}) = \frac{|F|}{|F|+T}$. Let $w' \neq w'' \neq w$. Since Pr is regular, $Pr(\{w, w'\}) > \frac{|F|}{|F|+T}$ and $Pr(\{w, w''\}) > \frac{|F|}{|F|+T}$

Now consider the belief set $B = \{p : Pr(p) \geq \frac{|F|}{|F|+T}\} \setminus \{\{w\}\}$. For all $p \in B$, $Pr(p) \geq \frac{|F|}{|F|+T}$ and so by the theorem above, B is neither weakly nor strongly W-dominated. But B is not multi-premise closed, for $\{w, w'\}, \{w, w''\} \in B$ but $\{w\} \notin B$. □

[52] Easwaran 2016, p. 827.

[53] Many thanks to Kenny Easwaran for suggesting this proof, which is much quicker than the earlier version I had.

Appendix 1.C Single-premise closure is not required for maximising expected accuracy

Proposition 26. *Single-premise closure is not necessary for maximising expected accuracy.*

Proof. Let $W = \{w_1, w_2, w_3\}, Pr(\{w_1, w_2\}) = 1$ and $Pr(\{w_1\}) = Pr(\{w_2\}) = \frac{1}{2}$. Let $T = |F|$ and $T > 0 > F$. Then the following belief B set maximises expected accuracy relative to Pr, for all $p \in B$ are such that $Pr(p) \geq t = \frac{|F|}{|F|+T}$ — equivalently, for all $p \in B$, $Pr(p) \geq \frac{|F|}{|F|+T}$ — and therefore B is neither weakly nor strongly W-dominated.

$$B = \{W, \{w_1, w_2\}, \{w_1\}\}$$

It is easy to check that for every $p \in B$, $Pr(p) \geq \frac{1}{2}$, and for no $p \notin B$, $Pr(p) > \frac{1}{2}$. But note that B violates single-premise closure, for $\{w_1, w_3\} \notin B$ even though $\{w_1\} \in B$. □

Because $Pr(\{w_1\}) = Pr(\{w_1, w_3\}) = \frac{1}{2}$, from the point of view of accuracy it makes no difference whether they are in B or not; the expected value is the same. And B is a belief set where single-premise closure fails even though B maximises expected accuracy with respect to Pr.

2 | Accuracy and higher-order belief

This chapter extends the investigation of accuracy coherence initiated in the previous chapter, with the aim of learning more about what requirements this kind of accuracy coherence places on higher-order beliefs: that is, beliefs about our beliefs. Specifically, we examine what the kinds of accuracy coherence we have discussed have to say about belief introspection principles, anti-expertise, and Moorean beliefs.

2.1 Introduction

This chapter investigates the relationship between the accuracy-theoretic notions of belief coherence discussed in the previous chapter, and various questions about higher-order beliefs — that is, what you believe about what you believe. Technically, we handle this by integrating the results of the previous chapter with doxastic logic frames and examining the conditions under which frames are such that every world in the frame represents an accuracy-coherent belief set. In effect, we can investigate the 'logic' of the kinds of accuracy-coherence we have discussed.

With that integration effected, we ask three sets of questions. First, are there accuracy-centred motivations for so-called belief introspection principles, according to which, whenever you have (lack) a belief about some proposition you believe that you have (lack) a belief about that proposition. Second, are there accuracy-centred reasons to avoid thinking of yourself as an anti-expert; as someone whose beliefs are — in certain respects — a poor guide to the truth. Thirdly, are there accuracy-centred reasons to avoid holding Moorean beliefs, like the belief that: p, but you don't believe p to be true.

The discussion in this chapter is more exploratory than the previous, and the epistemology of higher order belief that tentatively emerges has its attractions as

well as its deficits.[1] In the final part of the chapter, I look at a different way of investigating the questions described above — one that we might hope to secure a less tentative improvement on the former picture — and I explain why I'm cautious about employing this alternative framework.

Here is a roadmap. In section 2.2 I lay out some background technical resources which allow us to examine questions about higher-order beliefs. In section 2.3, we look at what this framework tells us about belief introspection principles. In section 2.4 we look at what it tells us about certain kinds of anti-expertise. In section 2.5, we look at Moorean beliefs. In section 2.6 we consider an alternative framework for thinking about these questions, though we express some reservations about investigating things this way. Section 2.7 concludes.

2.2 Accuracy coherence and doxastic logic frames

2.2.1 Formal background

Our first concern is the technical matter of integrating the notions of accuracy coherence we have already met with doxastic logic frames. This is the additional formal resource needed for us to represent the higher-order beliefs of agents. Our primary focus will be what we have previously called $\cap B$-coherence — for technical reasons to be introduced shortly, we will have to slightly rename this — though we will also be able to draw some conclusions about W-coherence. To make our later discussion tractable, we have to restate a number of results we have already seen in slightly different notation.

Consequently, we introduce a *pre-frame*, a triple $\langle W, T, F\rangle$. W is the same

[1] As we'll see later on, our approach has affinities with the investigation of credal deference principles in Dorst (2020), which is (at least in part) also concerned with issues of accuracy. See also Schoenfield (2016); Dorst, Levinstein, Salow, Husic, and Fitelson (2021) for accuracy-flavoured approaches to questions about deference.

finite set of mutually exclusive and jointly exhaustive possible worlds. T, F are numbers $\in \mathbb{R}$. We place one constraint on their values: that $T > 0 > F$. We again appeal to accuracy functions for belief sets:

Definition 1. *A^b is an accuracy function iff* $A^b(B, w) = |\{p \in B : w \in p\}|T + |\{p \in B : w \notin p\}|F$

Informally, then, for a given pre-frame, A^b assigns B a score of T for each proposition true at w, and F for each proposition false at w. The restriction that $T > 0 > F$ then encodes the assumption that true beliefs are epistemically good and false beliefs are epistemically bad. We motivated thinking about accuracy for outright beliefs this way in the previous chapter.

2.2.2 $\cap$-coherence

As we have already seen, for any belief set B, the proposition $\cap B$ is the set of worlds consistent with what the agent believes. At every world in $\cap B$, all the propositions in B are true. We can interpret $\cap B$ as the set of 'live' worlds according to a given B. We now make use of $\cap B$ in stating some dominance conditions. These are effectively the dominance conditions we met in the previous chapter, though as we will see before long, it is convenient to state them with a third variable.

Definition 41 (Strong $\cap$-dominance with respect to B)**.** *B'' strongly $\cap$-dominates B' with respect to B iff $A^b(B'', w) > A^b(B', w)$ at all $w \in \cap B$.*

Definition 42 (Weak $\cap$-dominance with respect to B)**.** *B'' weakly $\cap$-dominates B' with respect to B iff $A^b(B'', w) \geq A^b(B', w)$ at all $w \in \cap B$, and $A^b(B'', w) > A^b(B', w)$ at some $w \in \cap B$.*

Definition 43 (Weak $\cap$-dominance of B' with respect to B)**.** *B' is weakly $\cap$-dominated with respect to B iff there is some B'' such that B'' weakly $\cap$-dominates B' with respect to B.*

Definition 44 (Strong $\cap$-dominance of B' with respect to B)**.** *B' is strongly $\cap$-dominated with respect to B iff there is some B'' such that B'' strongly $\cap$-dominates B' with respect to B.*

I will occasionally omit the 'with respect to' construction when it is either clear from context which belief set is being referred to, or irrelevant. The above dominance conditions are combined quite naturally with the following normative claim:

> **$\cap$-coherence**: It's rationally prohibited for B to be either strongly or weakly $\cap$-dominated with respect to B.

The normative claim places a necessary condition upon epistemic rationality. It says that it should never be the case that B is $\cap$-dominated with respect to itself. I will not say much about why we should endorse $\cap$-coherence; this happened in the previous chapter. For very quick motivation we might note that $\cap$-coherence amounts to a sort of 'internal coherence' condition, that requires you to do as well as possible by your own lights, and that $\cap$-coherence is analogous to the popular credal accuracy principle *propriety*; see the previous chapter for fuller discussion.

Some further terminology:

Definition 45 ($\cap$-coherence of B' with respect to B)**.** *A belief set B' is $\cap$-coherent with respect to B just in case B' is neither strongly nor weakly $\cap$-dominated with respect to B.*

Lastly, just for reference in what follows, I state two further definitions.

Definition 46 (Consistency)**.** *B is consistent iff $\cap B \neq \emptyset$*

Definition 47 (Multi-premise closure)**.** *B is multi-premise closed with respect to W iff $\cup B \subseteq W$ and if $\cap B \subseteq p$, then $p \in B$.*

2.2.3 Necessary and sufficient conditions for $\cap$-coherence of B

We have shown before that the following holds:

Proposition 48. *For all pre-frames $\langle W, T, F\rangle$, for all B, if B is $\cap$-coherent with respect to itself, then B is consistent and multi-premise closed with respect to W.*

We repeat the proof, using our new notation.

Proposition 49. *For all pre-frames, for all B, if B is inconsistent, B is not $\cap$-coherent with respect to itself.*

Proof. If B is inconsistent, then $\cap B = \emptyset$. Then B is vacuously strongly $\cap$-dominated. □

Proposition 50. *For all pre-frames, for all B, for all $w \in \cap B$, $A^b(B, w) = |B|T$.*

Proof. By definition 1, $A^b(B, w) = |\{p \in B : w \in p\}|T + |\{p \in B : w \notin p\}|F$. At $w \in \cap B$, $A^b(B, w) = |\{p \in B : w \in p\}|T = |B|T$. □

Proposition 51. *For all pre-frames $\langle W, T, F\rangle$, for all B, if B isn't multi-premise closed with respect to W, then B is not $\cap$-coherent with respect to itself.*

Proof. Suppose B is not consistent. Then by proposition 49, B is not $\cap$-coherent with respect to itself.

Next, suppose B is consistent, but not multi-premise closed with respect to W. Then by definition of multi-premise closure, there is some p such that $\cap B \subseteq p$ and $p \notin B$. But then by proposition 50, $A^b(B \cup \{p\}, w) > A^b(B, w)$ at all $w \in \cap B$. Therefore B is strongly $\cap$-dominated with respect to B, and B is not $\cap$-coherent. □

Propositions 49 and 51 jointly establish proposition 48.

In this section we provide necessary *and* sufficient conditions for some B to be $\cap$-coherent with respect to itself (that is to say, that B does not violate $\cap$-coherence).

The expected accuracy EA^b of B relative to some Pr is defined as follows:

Definition 52 (Expected Accuracy)**.**

$$EA^b_{Pr}(B) = \sum_{w \in W} Pr(\{w\}) A^b(B, w)$$

When it is clear from context which probability function the expected accuracy is relative to, I may omit the index.

Definition 53 (Maximising Expected Accuracy)**.** *B maximises expected accuracy with respect to Pr iff for all B', $EA^b_{Pr}(B) \geq EA^b_{Pr}(B')$.*

We also recall the following definition.

Definition 27 (Regularity)**.** *A probability function Pr is regular iff $Pr(\{w\}) > 0$ for all $w \in W$.*

As we saw in the previous chapter, Easwaran shows the following (we re-state this in the setting of pre-frames):

Lemma 54 (Easwaran's Theorem 1)**.** *For all pre-frames $\langle W, T, F \rangle$, for all B, if B maximises expected accuracy with respect to some regular Pr, then there is no B' such that $A^b(B', w) \geq A^b(B, w)$ at all $w \in W$ and $A^b(B', w) > A^b(B, w)$ at some $w \in W$.*[2]

We can generalise Easwaran's result in one respect. Call a probability function *irregular* iff it is not regular. Easwaran does not consider irregular probability

[2]Easwaran (2016, pp. 827-29).

functions. This is because, as we have already seen, he is interested in a different kind of belief dominance and coherence.[3]

We now show that if B maximises expected accuracy with respect to Pr, then no B' outscores B at all $w \in \{w : Pr(\{w\}) > 0\}$. This holds even in the case where $\{w : Pr(\{w\}) > 0\} \neq W$ i.e. where Pr is irregular.

Proposition 55. *For every pre-frame, for all B, if B maximises expected accuracy with respect to any Pr then there is no B' such that $A^b(B', w) \geq A^b(B, w)$ at all $w \in \{w : Pr(\{w\}) > 0\}$ and $A^b(B', w) > A^b(B, w)$ at some $w \in \{w : Pr(\{w\}) > 0\}$.*

Proof. Suppose B maximises expected accuracy with respect to Pr. Then:

$$\sum_{w \in \{w : Pr(\{w\}) > 0\}} Pr(\{w\}) A^b(B, w) \geq \sum_{w \in \{w : Pr(\{w\}) > 0\}} Pr(\{w\}) A^b(B', w)$$

for all B'; since those $w \notin \{w : Pr(\{w\}) > 0\}$ receive probability zero and are therefore not counted by the summation.

Suppose moreover for reductio that there is some B' such that, for all $w \in \{w : Pr(\{w\}) > 0\}$, $A^b(B', w) \geq A^b(B, w)$ and for some $w' \in \{w : Pr(\{w\}) > 0\}$, $A^b(B', w') > A^b(B, w')$. So for every term in the summation, $Pr(\{w\}) A^b(B', w) \geq Pr(\{w\}) A^b(B, w)$, and for some term $Pr(\{w'\}) A^b(B', w') > Pr(\{w'\}) A^b(B, w')$.

Rearranging, the expected accuracy of $B' \geq$

$$Pr(\{w'\}) A^b(B', w') \; + \sum_{w \in \{w : Pr(\{w\}) > 0\} \setminus \{w'\}} Pr(\{w\}) A^b(B', w)$$

and the expected accuracy of $B =$

[3] We will discuss this alternative notion again in the appendix.

$$Pr(\{w'\})A^b(B, w') \; + \sum_{w \in \{w : Pr(\{w\}) > 0\} \setminus \{w'\}} Pr(\{w\})A^b(B, w)$$

Since $Pr(\{w'\})A^b(B', w') > Pr(\{w'\})A^b(B, w')$ and $Pr(\{w\})A^b(B', w) \geq Pr(\{w\})A^b(B, w)$ for all $w \in \{w : Pr(\{w\}) > 0\} \setminus \{w'\}$, then $EA^b(B') > EA^b(B)$. But then B would not maximise expected accuracy relative to Pr, contrary to supposition.

□

This generalisation of Easwaran's result is helpful. To see why, we need to inspect our differing notions of $\cap$-dominance more closely:

Definition 56 (Vacuous strong $\cap$-dominance)**.** *B'' vacuously strongly $\cap$-dominates B' with respect to B iff B'' strongly $\cap$-dominates B' and $\cap B = \emptyset$.*

Definition 57 (Non-vacuous strong $\cap$-dominance)**.** *B'' non-vacuously strongly $\cap$-dominates B' with respect to B iff B'' strongly $\cap$-dominates B' with respect to B and $\cap B \neq \emptyset$.*

There is no 'vacuous' weak $\cap$-dominance because of the existential quantifier in its definition.

We can now argue as follows:

Proposition 58. *For all pre-frames $\langle W, T, F\rangle$, for all B, if $\cap B = \{w : Pr(\{w\}) > 0\}$, and B is multi-premise closed with respect to W, then if B maximises expected accuracy with respect to Pr, B is $\cap$-coherent with respect to B.*

Proof. Suppose $\cap B = \{w : Pr(\{w\}) > 0\}$, B is multi-premise closed with respect to W, and B maximises expected accuracy with respect to Pr. Therefore by proposition 55, B is not non-vacuously strongly $\cap$-dominated with respect to B, and not weakly $\cap$-dominated with respect to B. And since $\cap B = \{w :$

$Pr\{w\} > 0\}$, $Pr(\cap B) = 1, \cap B \neq \emptyset$. For in general, $Pr(\emptyset) = 0$. Therefore B is not vacuously strongly $\cap$-dominated with respect to B. Therefore B is $\cap$-coherent with respect to itself. □

We have shown that when Pr assigns some probability to all worlds in $\cap B$, and no probability to worlds not in $\cap B$, then if B is (consistent and) multi-premise closed with respect to W, and maximises expected accuracy with respect to Pr, then B is $\cap$-coherent with respect to itself. In what follows we will be interested in a particular probability function that assigns some probability to all $w \in \cap B$.

Definition 59. *For all* B, $Pr_B(\{w\}) =_{def} \begin{cases} \frac{1}{|\cap B|} & w \in \cap B \\ 0 & w \notin \cap B \end{cases}$

From proposition 58 it follows that if B is multi-premise closed and maximises expected accuracy with respect to Pr_B, it is $\cap$-coherent with respect to itself. We now turn our attention to the other direction: if B is $\cap$-coherent with respect to itself then B maximises expected accuracy with respect to Pr_B.

First we recall another result, from Easwaran (2016, pp. 827-29). We referred to a version of this result from Kevin Dorst in the previous chapter. It says the following:

Proposition 60 (Easwaran's Theorem 2)**.** *For every pre-frame, for all* B*, for all* Pr*,* B *maximises expected accuracy with respect to* Pr *iff* $Pr(p) \geq \frac{|F|}{|F|+T}$ *for all* $p \in B$ *and* $Pr(p) > \frac{|F|}{|F|+T}$ *for no* $p \notin B$*.*

We now need the following lemma.

Lemma 61. *For all pre-frames, for all* B*, for all* Pr*: suppose that for all* p *and* q*, if* $Pr(p) = Pr(q)$ *then* $p \in B$ *iff* $q \in B$*. Then for all* w *and* w'*, if* $Pr(\{w\}) = Pr(\{w'\})$*, then* $A^b(B, w) = A^b(B, w')$*.*

Proof. Let ϕ be the permutation of W such that $\phi(w) = w', \phi(w') = w$, and $\phi(w'') = w''$ for all other w''. Let $\phi*$ be the permutation of $P(W)$ such that $\phi*(p) = \{\phi(w) : w \in p\}$ for all p. Since $Pr(\{w\}) = Pr(\{w'\})$, $Pr(\phi*(p)) = Pr(p)$ for all p. For

$$Pr(\phi*(p)) = \sum_{w \in p} Pr(\{\phi(w)\}) = \sum_{w \in p} Pr(\{w\}) = Pr(p)$$

Hence, by the condition on B, $\phi*(p) \in B$ iff $p \in B$ for all p. Hence $\phi*$ is a bijection between the members of B true at w and those true at w', and between the members of of B false at w and those false at w'. Thus $A^b(B, w) = A^b(B, w')$. □

Proposition 62. *For all pre-frames, for all B, if B is $\cap$-coherent with respect to itself then B maximises expected accuracy relative to Pr_B.*

Proof. Suppose B is $\cap$-coherent with respect to B. Then by proposition 48, B is consistent. Since W has finitely many members, and both T and F are finite in value, there is some belief set, $B' = \{p : Pr(p) \geq \frac{|F|}{|F|+T}\}$ which by proposition 60, maximises expected accuracy with respect to Pr_B. Suppose that $B' \neq B$. By lemma 61, since $Pr_B(\{w\}) = Pr_B(\{w'\})$ for all $w, w' \in \cap B$, then $A^b(B', w) = A^b(B', w')$ at all $w, w' \in \cap B$. Since B' maximises expected accuracy with respect to Pr_B, then $EA^b_{Pr_B}(B') \geq EA^b_{Pr_B}(B)$. There are now two cases to consider.

1. In this case, $EA^b_{Pr_B}(B') > EA^b_{Pr_B}(B)$. Since by lemma 61 B' scores the same at all worlds in $\cap B$, then to have greater expected accuracy than B, it must outscore B at every world in $\cap B$. For if it scores equal or less than B at any world in $\cap B$, it scores equal or less than B at all worlds in $\cap B$ and thereby cannot have greater accuracy in expectation relative to Pr_B. In this case, B' strongly

$\cap$-dominates B with respect to B, meaning B is not $\cap$-coherent with respect to itself — contrary to our supposition.

2. In case 2, $EA^b_{Pr_B}(B') = EA^b_{Pr_B}(B)$. Since B' maximises expected accuracy with respect to Pr, and B has identical expected accuracy to B', then B also maximises expected accuracy with respect to Pr.

□

Putting together propositions 58 and 62, we get the following quite powerful proposition:

Proposition 63. *For all pre-frames $\langle W, T, F\rangle$, any belief set B is $\cap$-coherent with respect to itself iff B is consistent B is multi-premise closed with respect to W, and B maximises expected accuracy relative to Pr_B.*

Using this we can draw some other important consequences.

Proposition 64. *For all pre-frames $\langle W, T, F\rangle$, for all B that are consistent and multi-premise closed with respect to W, B maximises expected accuracy with respect to Pr_B iff $\frac{|\cap B|-1}{|\cap B|} \leq \frac{|F|}{|F|+T}$.*

Proof.

$\Rightarrow$ Suppose B maximises EA with respect to Pr_B. By proposition 60, since B maximises EA, for all $p \notin B$, $Pr_B(p) \leq \frac{|F|}{|F|+T}$. Since all worlds in $\cap B$ receive equal probability, the propositions $p \subset \cap B$ such that $|p| = |\cap B|-1$, i.e. those propositions with $\cap B - 1$ worlds, have probability at least as great as any propositions $\notin B$. By assumption, for $w \in \cap B$, $Pr_B(\{w\}) = \frac{1}{|\cap B|}$. Therefore $\frac{|\cap B|-1}{\cap B} \leq \frac{|F|}{|F|+T}$.

$\Leftarrow$ Suppose $\frac{|\cap B|-1}{|\cap B|} \leq \frac{|F|}{|F|+T}$.

We show that for all $p \in B$, $Pr_B(p) \geq \frac{|F|}{|F|+T}$, and for no $p \notin B$, $Pr_B(p) > \frac{|F|}{|F|+T}$. It follows by Easwaran's Theorem 2 (proposition 60, above) that B maximises expected accuracy with respect to Pr_B.

By assumption, $Pr_B(\cap B) = 1$; since B is multi-premise closed with respect to W, it follows that for all $p \in B$, $Pr_B(p) = 1 \geq \frac{|F|}{|F|+T}$. It remains to show that for no $p \notin B$, $Pr_B(p) > \frac{|F|}{|F|+T}$.

Since for all $w \in \cap B$, $Pr_B(\{w\}) = \frac{1}{|\cap B|}$, then for all $X \subset \cap B$, $|X| \leq |\cap B| - 1$. Therefore $Pr_B(X) \leq \frac{|\cap B|-1}{\cap B}$, which by assumption is less than or equal to $\frac{|F|}{|F|+T}$. Since X has at least as great probability as any $p \notin B$, B maximises expected accuracy with respect to Pr_B.

□

Proposition 65. *For all pre-frames $\langle W, T, F\rangle$, for all B that are consistent and multi-premise closed with respect to W, $|\cap B| > 1$ iff there is some pre-frame $\langle W, T', F'\rangle$ where B does not maximise expected accuracy with respect to Pr_B.*

Proof.

⇒ For arbitrary pre-frame $\langle W, T, F\rangle$, suppose B is consistent, multi-premise closed with respect to W and that $|\cap B| > 1$. Then $\frac{|\cap B|-1}{|\cap B|} \geq \frac{1}{2}$. Let $T' > |F'|$ so that $\frac{|F'|}{|F'|+T'} < \frac{1}{2}$. So by proposition 64, in the pre-frame $\langle W, T', F'\rangle$, B does not maximise expected accuracy with respect to Pr_B.

⇐ Contrapose the converse: for all pre-frames $\langle W, T, F\rangle$ for all B that are consistent, multi-premise closed with respect to W, if $|\cap B| = 1$, then there's no pre-frame $\langle W, T', F'\rangle$ where B does not maximise expected accuracy with respect to Pr_B. For suppose B is consistent, multi-premise closed with respect to W and $|\cap B| = 1$. Then for all $w \in \cap B$, $Pr_B(\{w\}) = \frac{1}{|\cap B|} = \frac{1}{1} = 1$. Then all and only $p \in B$ are such that $Pr_B(p) \geq \frac{|F|}{|F|+T}$ for

every proposition not in B receives probability zero. So by proposition 60, B maximises expected accuracy with respect to Pr_B.

□

Because of this, we can conclude that:

Proposition 66. *For all pre-frames $\langle W, T, F\rangle$, for all belief sets, B is $\cap$-coherent with respect to itself relative to all pre-frames iff B is multi-premise closed with respect to W and $|\cap B| = 1$.*

Proof. Trivially follows from proof of proposition 65. □

Relatedly:

Proposition 67. *For all pre-frames $\langle W, T, F\rangle$ where $T > |F|$, any B is $\cap$-coherent with respect to itself iff $|\cap B| = 1$.*

Proof. Since $T > |F|$, it follows that $\frac{|F|}{|F|+T} < \frac{1}{2}$. Suppose $|\cap B| \geq 2$. Then $\frac{|\cap B|-1}{|\cap B|} \geq \frac{1}{2}$, and therefore by lemma 65, B is not $\cap$-coherent with respect to itself. By proposition 66, we know that if B is consistent and multi-premise closed with respect to W, and $|\cap B| = 1$, then B is $\cap$-coherent with respect to itself. It follows that in pre-frames where $T > |F|$, the only B which are $\cap$-coherent with respect to themselves are those where $|\cap B| = 1$. □

Another interesting related consequence is the following:

Proposition 68. *For all pre-frames $\langle W, T, F\rangle$, if $\{W\}$ is $\cap$-coherent with respect to itself, then $|F| \geq (|W| - 1)T$*

Proof. By proposition 63, $\cap$-coherence of $\{W\}$ with respect to itself implies it maximises expected accuracy relative to Pr_B. Since $\cap B = W$, this is just to say that $Pr(W) = 1$. Therefore by proposition 64, for $\{W\}$ to maximise expected

accuracy relative to Pr must be the case that $\frac{|W|-1}{|W|} \leq \frac{|F|}{|F|+T}$. Suppose for reductio that $F < (|W|-1)T$. Then $\frac{|F|}{|F|+T} < \frac{(|W|-1)T}{(|W|-1)T+T} = \frac{(|W|-1)T}{|W|T} = \frac{|W|-1}{|W|}$, so we reject our supposition that $F < (|W|-1)T$. □

As saw in the previous chapter (though not within the setting of what we are calling 'pre-frames'), suspended judgment is naturally modelled as believing neither a proposition nor its negation. So in effect, this result explains what a pre-frame must be like for it to be permissible (by the lights of ∩-coherence) to suspend judgment on absolutely everything non-trivial.

2.2.4 Doxastic logic

Now that we've put in place a better understanding of the necessary and sufficient conditions for ∩-coherence of some B with respect to itself, it's time to enrich our framework. We first introduce a Kripke frame, which is a pair $\langle W, R\rangle$. W is just the set of worlds described above. R is a dyadic accessibility relation over W. It is R that lets us represent propositions about what is believed. When Rww', we say that w' is doxastically possible relative to w. We define the proposition $R(w)$ as $\{w' : Rww'\}$; in other words $R(w)$ is the set of doxastic possibilities at w. For a world w' to be in $R(w)$ just means that, for all the agent believes, she is in w'. An agent believes some proposition p — represented as $\mathbf{B}p$ — at w just in case $R(w) \subseteq p$; just in case p is true at every doxastically accessible world. That is, $\mathbf{B}p$ is true at w just in case $R(w) \subseteq p$. This means that $R(w)$ is the strongest thing believed at w — a fact which will be useful in what follows. This also means that on Kripke frames, the represented belief sets are automatically multi-premise closed with respect to W.

With this additional resource, our objects of study are therefore *accuracy frames*, quadruples $\langle W, T, F, R\rangle$; i.e. the pre-frames we have previously discussed, but enriched with an accessibility relation.

We can use the notion of $\cap$-coherence of B with respect to itself that we were working with before to impose $\cap$-coherence on the accuracy frames. We do so by introducing a notion of $\cap$-coherence for accuracy frames. First, let us define $B(w)$ as the set of believed propositions at w in some accuracy (or Kripke) frame. Equivalently, $B(w)$ is the set of all supersets of $R(w)$. Note, therefore, that given standard set theoretical assumptions, then for any given w, $B(w) \neq R(w)$.

Proposition 69. *Given ZF, for all w, $R(w) \neq B(w)$*

Proof. Suppose that for some w, $R(w) = B(w)$. Trivially, $W \in B(w)$, for any $B(w)$ is automatically multi-premise closed with respect to W. Therefore $W \in R(w)$. But $R(w) \subseteq W$, meaning everything in $R(w)$ is in W. But that is just to say, W is a member of itself, which is not possible in ZF. □

We can now define frame $\cap$-coherence.

Definition 70 (Frame $\cap$-coherence)**.** *An accuracy frame $\langle W, R, T, F\rangle$ is frame $\cap$-coherent iff $\forall w$ $B(w)$ is $\cap$-coherent with respect to $B(w)$.*

I'll be a bit loose with the 'accuracy frame' terminology; it should always be clear from context whether I'm taking 'frame' to refer to an accuracy frame (the quadruple described above) or a frame (a pair $\langle W, R\rangle$).

2.2.5 Accuracy Frame $\cap$-coherence

What conditions are necessary and sufficient for an accuracy frame $\langle W, R, T, F\rangle$ to be frame $\cap$-coherent? To see one necessary condition, let us define a serial relation and consequently, a serial accuracy frame.

Definition 71. *A relation R is serial iff for all w there is some w' such that Rww'.*

Definition 72. *An accuracy frame $\langle W, R, T, F\rangle$ is serial iff R is serial.*

We can now show that:

Proposition 73. *If $\langle W, R, T, F\rangle$ is frame $\cap$-coherent, then $\langle W, R, T, F\rangle$ is serial.*

Proof. Suppose $\langle W, R, T, F\rangle$ is not serial. Then R is not serial. Then there is some $w \in W$ such that $R(w) = \emptyset$. Then $B(w)$ is not consistent. By proposition 48, $B(w)$ is not $\cap$-coherent with respect to itself. So $\langle W, R, T, F\rangle$ is not accuracy frame $\cap$-coherent. □

Before we turn our attention to more specific questions about higher-order belief, there are some further results about these accuracy frames that it will be useful to prove.

Definition 74 (Robustness)**.** *An accuracy frame $\langle W, R, T, F\rangle$ is robust iff for all T', F' such that $T' > 0 > F'$, $\langle W, R, T', F'\rangle$ is frame $\cap$-coherent.*

One way of thinking about robustness is as follows: if an accuracy frame is robust, then however we vary the values of T and F (while respecting the minimal requirement that $T > 0 > F$) we can never end up at a world where our beliefs are not $\cap$-coherent. Robustness provide a kind of insulation from variation in the relative goodness and badness of truth and falsity.

Proposition 75. *An accuracy frame is robust iff R is a one-to-one relation.*

Proof.

$\Rightarrow$ An accuracy frame $\langle W, R, T, F\rangle$ is frame $\cap$-coherent iff for all $w \in W$, $B(w)$ is $\cap$-coherent with respect to itself. By proposition 65, if $|\cap B(w)| > 1$ then there are some T, F such that $B(w)$ is not $\cap$-coherent with respect to itself. Therefore if $B(w)$ is $\cap$-coherent with respect to itself relative to all values for T, F, then $|\cap B(w)| = 1$. This means that for all w, there is only one world w' such that $R(w) = w'$.

$\Leftarrow$ Suppose that for each w, there is a unique world w' such that $R(w) = \{w'\}$. Then for each $B(w)$, $|\cap B(w)| = 1$. Therefore by proposition 65, each $B(w)$ is $\cap$-coherent with respect to itself relative to all T, F. So for all T', F', $\langle W, R, T, F\rangle$ is frame $\cap$-coherent. So $\langle W, R, T, F\rangle$ is robust.

□

Robustness therefore corresponds to the coincidence of (doxastic) possibility and necessity. Moreover, robust frames represent the opinions of a maximally opinionated, consistent believer. Maximally opinionated in the sense that, for each $B(w)$ and every proposition p, exactly one of p and $\neg p$ is believed; consistent in the sense that all the believed propositions are true at some world — namely, $\cap B(w)$.

Definition 76 (Reflexivity)**.** *R is reflexive iff $\forall w\ Rww$.*

Definition 77. *An accuracy frame $\langle W, R, T, F\rangle$ is reflexive iff R is reflexive.*

Incidentally, this in turn means that for robust frames, whenever R is reflexive, then R is identity.

Proposition 78. *If $\langle W, R, T, F\rangle$ is robust, then if R is reflexive, R is identity.*

Proof. Trivially follows from proposition 65 and the definition of reflexivity. □

I want to make one further observation about proposition 64, which is that it also lets us state necessary and sufficient conditions for frame $\cap$-coherence of a frame.

Proposition 79. *A serial accuracy frame $\langle W, R, T, F\rangle$ is frame $\cap$-coherent iff for all w, $\frac{|\cap B(w)|-1}{|\cap B(w)|} \leq \frac{|F|}{|F|+T}$*

Proof.

$\Rightarrow$ If $\langle W, R, T, F\rangle$ is frame $\cap$-coherent then for all w, $B(w)$ is $\cap$-coherent with respect to itself. By proposition 63, $B(w)$ therefore maximises expected accuracy relative to $Pr_{B(w)}$. Therefore by proposition 64, $\frac{|\cap B(w)|-1}{\cap B(w)} \leq \frac{|F|}{|F|+T}$.

$\Leftarrow$ If for all w, $\frac{|\cap B(w)|-1}{\cap B(w)} \leq \frac{|F|}{|F|+T}$, then by the proof of proposition 64, for all w, $B(w)$ maximises expected accuracy with respect to $Pr_{B(w)}$. So by proposition 63, for all w, $B(w)$ is $\cap$-coherent with respect to itself, which is just the definition of frame $\cap$-coherence.

□

This in turn lets us show that frame $\cap$-coherence is closed under finite disjoint union, assuming that the frames being unioned share the same values for T and F.

Proposition 80. *If some accuracy frames $\langle W_1, R_1, T, F\rangle, ..., \langle W_n, R_n, T, F\rangle$, are frame $\cap$-coherent, and W_1 through W_n are pairwise disjoint, then $\langle W \cup ... \cup W_n, R \cup ... \cup R_n, T, F\rangle$ is frame $\cap$-coherent.*

Proof. We show that if $\langle W, R, T, F\rangle$ and $\langle W', R', T, F\rangle$ are frame $\cap$-coherent, and W and W' are disjoint, then $\langle W \cup W', R \cup R', T, F\rangle$ is frame $\cap$-coherent. Since W and W' are disjoint, R is a relation on W and R' is a relation on W', then for all $w \in W$, $R(w) = R \cup R'(w)$, and for all $w' \in W'$, $R'(w') = R \cup R'(w')$; we abbreviate this fact as $(*)$.

Since $\langle W, R, T, F\rangle$ and $\langle W', R', T, F'\rangle$ are both frame $\cap$-coherent, by proposition 79 it follows that for all $w \in W$, $\frac{|\cap B(w)|-1}{\cap B(w)} \leq \frac{|F|}{|F|+T}$, and likewise, for all $w' \in W'$, $\frac{|\cap B(w')|-1}{\cap B(w')} \leq \frac{|F|}{|F|+T}$. Recall that in general, $R(w) = \cap B(w)$, and so for all $w \in W \cup W'$, $R \cup R'(w) = \cap B(w)$. Then by $(*)$, for all $w'' \in W \cup W'$, $\frac{|\cap B(w'')|-1}{\cap B(w'')} \leq \frac{|F|}{|F|+T}$. Therefore $\langle W \cup W', R \cup R', T, F\rangle$ is frame $\cap$-coherent. The remainder follows inductively. □

2.3 Belief introspection

These technical investigations put us in a position to ask more philosophical questions. The first thing we will investigate is whether frame coherence requires belief introspection principles to hold. In other words, we ask whether considerations of accuracy require certain kinds of belief introspection. To state these principles, we need to define the following properties.

Definition 81. *A relation R is transitive iff if Rww' and $Rw'w''$ then Rww''.*

Definition 82. *An accuracy frame $\langle W, R, T, F\rangle$ is transitive iff R is transitive.*

Definition 83. *A relation R is euclidean iff if Rww' and Rww'' then $Rw'w''$.*

Definition 84. *An accuracy frame $\langle W, R, T, F\rangle$ is euclidean iff R is euclidean.*

If an accuracy frame is transitive, then for all propositions p that the agent believes, the agent *believes* that they believe p. If an accuracy frame is euclidean, then for all propositions p that the agent doesn't believe, the agent *believes* that they don't believe p. The properties therefore amount to so-called 'introspection' principles, principles that place general, structural constraints upon the beliefs an agent has about their own beliefs.

There are some trivial 'introspective' beliefs that agents are guaranteed to have just in virtue of the logical structure of the frame. For instance, for all w, $R(w) \subseteq W$, so $\mathbf{B}W$ is true at every w, and therefore $\mathbf{BB}W$ is true at every w — and so on. Similarly, since frame $\cap$-coherent frames are serial, at no w is $R(w) \subseteq \emptyset$. Therefore at every w, $\neg\mathbf{B}\emptyset$ is true. Accordingly, at every world, $\mathbf{B}\neg\mathbf{B}\emptyset$ is true.[4]

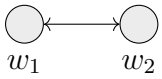
Some less trivial results are the following:

[4]It's probably technically sloppy of me to use '$\neg$' here, as that's standardly a sentential operator and $\mathbf{B}\emptyset$, as introduced, is a proposition, i.e. a set of worlds. I use this, rather than '$W\backslash$', because I've found that despite the technical infelicity, it facilitates readability. But do note that this is loose speak on my part.

Proposition 85. *There are no values of T and F such that, all frame $\cap$-coherent accuracy frames with those values are transitive.*

Proof. We describe an accuracy frame $\langle W, R, T, F\rangle$ that is robust but not transitive. It follows by definition of robustness that for all T', F' such that $T > 0 > F$, $\langle W, R, T', F'\rangle$ is frame $\cap$-coherent, and that there are no T, F such that all frames with those values are transitive.

Let $W = \{w_1, w_2\}$.



By proposition 75 this is a $\cap$-coherent frame relative to any T, F, since for all w, $|\cap B(w)| = 1$. However R is not transitive, since Rw_1w_2, Rw_2w_1, but it's not the case that Rw_1w_1.

□

Proposition 86. *There are no values of T and F such that, all frame $\cap$-coherent accuracy frames with those values are euclidean.*

Proof. We appeal again to the frame described in the previous proof. The frame is robust but not euclidean since Rw_1w_2, Rw_1w_2, but it is not the case that Rw_2w_2. The remainder of the proof proceeds as before. □

These show that for any values of T and F, including highly plausible ones, transitivity and the euclidean property are not in general required for frame $\cap$-coherence. In terms of the doxastic logic we are interested in, this means that there is no weighting of truth and falsity such that the avoidance of accuracy dominance requires believing that one believes some proposition, or believing that one doesn't believe some proposition. A natural way to interpret this result is as showing that there is no accuracy-centred motivation for belief introspection.

There is a related question we can ask. It is well known that when the doxastic logic we have been working with is interpreted epistemically — that is, when the worlds accessible to you are those consistent with what you know, rather than what you believe — the transitive and euclidean property are highly controversial. In this case, the negative introspection principle amounts to the claim that whenever you fail to know something, you know that you fail to know it. The positive introspection principle amounts to the claim that whenever you know something, you know that you know it. We will encounter the controversy over these principles in chapters four and five.

It may be tempting to think that the corresponding principles for belief are more plausible than the principles for knowledge. I will quickly gesture at why we might think this. Roughly, whether you know something is not something you have much control over — in an important sense, knowledge requires the co-operation of the world. Belief, on the other hand, intuitively requires no such co-operation. You can look out the window and believe it is raining, even though, if in an evil demon has conjured illusory raindrops, knowledge remains out of your grasp. Pursuing this line further, we might think that believing that you (don't) believe something *is* something that is easier to do than knowing that you (don't) know something. The latter again requires the co-operation of the world in a way that, intuitively, the former does not. Accordingly, we might think that the principle that if you believe something, you believe that you believe it, and the principle that if you don't believe something, you believe that you don't believe it, are plausible.[5] Just as we asked whether transitivity and the euclidean property fell out of any particular values of T and F, then, we can ask the converse. If we suppose a frame coherent accuracy frame is transitive, or euclidean, can we infer anything about the values of T and F? In particular, do those values come out looking

[5]In fact, I will express some reservations about this, too, in section 2.6.

(im)plausible in such a way that we can make judgments about whether the belief introspection principles are compatible with the framework we are working in?[6]

Proposition 87. *There exists an accuracy frame $\langle W, R, T, F\rangle$ such that $\langle W, R, T, F\rangle$ is transitive, $\cap$-coherent and for all T', F', $\langle W, R, T', F'\rangle$ is $\cap$-coherent.*

Proof. Let R be identity. Then R is transitive. Moreover, R is one-to-one and by proposition 75, $\langle W, R, T, F\rangle$ is robust. So for all T', F', $\langle W, R, T', F'\rangle$ is frame $\cap$-coherent. □

Proposition 88. *There exists an accuracy frame $\langle W, R, T, F\rangle$ such that $\langle W, R, T, F\rangle$ is euclidean, $\cap$-coherent and for all T', F', $\langle W, R, T, F\rangle$ is $\cap$-coherent.*

Proof. Let R be identity. Then R is euclidean. Moreover, R is one-to-one and by proposition 75, $\langle W, R, T, F\rangle$ is robust. So for all T', F', $\langle W, R, T', F'\rangle$ is frame $\cap$-coherent. □

What this means is that neither transitivity nor the euclidean property require any particular values for T and F. We cannot infer anything about the relative goodness and badness of false beliefs by assuming the introspection principles. These frame properties float free of the accuracy functions in both directions.

2.4 Anti-expertise

There are other interesting questions about the relationship between accuracy and higher-order beliefs. The accuracy frame described in our proof of proposition 85 raises a number of issues. Specifically, it is interesting because it describes a highly misleading learning situation: if you end up at w_1 then you will think you are at w_2 and vice versa.

[6] Cf. also Caie (2013, p. 531) in the credal case.

In terminology borrowed from Roy Sorensen, at each world, agents in frames like these are 'commissive anti-experts' (the significance of 'commissive' will become clear later). There is a proposition about which their beliefs are guaranteed to be wrong: if the proposition is false, they believe it true, and if it is true, they believe its negation.[7] For this reason, the frame considered above bears an obvious resemblance to the 'paradoxical' probabilistic frames considered in Williamson (2000, pp. 230-37), and to the 'sycophants' model in Dorst (2020, p. 601); in such frames, you receive 'anti-evidence', where the more likely it is you are at a world, the less likely it is that your evidence at that world tells you as much.

There is an uninteresting sense in which anti-expertise of this sort comes cheap. This is because $\cap$-coherence doesn't place any sort of factivity requirement upon belief. So it's consistent with $\cap$-coherence that you falsely believe p. But then by disjunction introduction, it's true that you either falsely believe p, or falsely believe $\neg p$ — and consequently, are a commissive anti-expert about p. What's significant about the frame above is that you're *guaranteed* to be a commissive anti-expert about p. Wherever you end up in the frame, you will end up an anti-expert, and end up believing as much.

It's noteworthy that $\cap$-coherence doesn't by itself seem impose enough constraints on the frames to block this kind of anti-expertise, as we saw in our proof of proposition 85. To repeat the lesson of our investigation into the belief introspection principles, the logic of $\cap$-coherence is too weak. However there is a way of imposing stricter constraints, at one remove. To see why, observe that at least in the frame we're presently considering, there's an important sense in which the agent can realise they are guaranteed to be made badly off when it comes to ac-

[7]This is to slightly misstate Sorensen's characterisation; for him "Someone is a commissive anti-expert about p if and only if his belief that p is strong evidence that $\neg p$ and his belief that $\neg p$ is strong evidence that p" (Sorensen, 1988, p. 392). For the sake of simplicity we assume more straightforwardly that commissive anti-expertise means beliefs are *guaranteed* their falsity; we follow Caie (2013, p. 530, fn. 2) in this usage.

curacy. First, note that relative to $B(w_1)$, you are $\cap$-coherent and (by your own lights) doing great, and likewise for $B(w_2)$. Upon stepping back from the frame, though, you can reason as follows: $A^b(B(w_1), w_1) = A^b(B(w_2), w_2) = T + F$. Whereas $A^b(\{W\}, w_1) = A^b(\{W\}, w_2) = T$. So when faced with a learning situation like the one described, suspending judgment on absolutely everything bar the tautology guarantees more accurate beliefs than following your evidence and ending up at either of w_1 or w_2. For the accuracy-minded thinker, you can know from the outset that such frames will make you badly off.

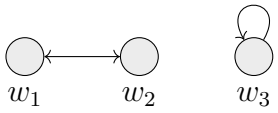
Just as we required when introducing this notion of accuracy coherence into the frames in the first place that every world represented a $\cap$-coherent belief set, it might be thought that there are also accuracy-centred reasons for requiring that the frames don't have the property that you can reason a priori that you can do better than any of the belief sets encoded in the frame. If we do impose this requirement, we can rule out this particular anti-expertise frame at one remove.

Let us make the story I sketched above more precise.

Definition 89. *A commissive anti-expertise proposition is a proposition of the form* $((p \cap \boldsymbol{B}\neg p) \cup (\neg p \cap \boldsymbol{B}p))$.

Definition 90. *An agent is a* commissive *anti-expert with respect to* p *iff* $((p \cap \boldsymbol{B}\neg p) \cup) \neg p \cap \boldsymbol{B}p))$.

The frame we have been considering is one in which you are a commissive anti-expert with respect to $\{w_1\}$ (and, moreover, $\{w_2\}$). In this frame, the agent is guaranteed to be a commissive anti-expert with respect to p. It's worth noting that it isn't generally the case that, if a frame allows that you might end up believing you're a commissive anti-expert, you're guaranteed to end up a commissive anti-expert. Consider the following frame:



At the first two worlds, you're a commissive anti-expert with respect to (for instance) $\{w_1\}$, and you believe as much. But that isn't the case at w_3, where you believe you aren't at w_1, and moreover, you aren't.

The second property we are interested in corresponds to the agent's being able to calculate that some other belief set is guaranteed to be at least as good, and possibly better, than any of belief sets captured in the frame.

Definition 91 (Self-effacing frame)**.** *An accuracy frame $\langle W, R, T, F\rangle$ is self-effacing iff there is some B such that for all w $A^b(B, w) \geq A^b(B(w), w)$ and for some w', $A^b(B, w') > A^b(B(w'), w')$.*

Definition 92. *$\langle W, R, T, F\rangle$ is self-effacing in virtue of B iff $\langle W, R, T, F\rangle$ is self-effacing and $A^b(B, w) \geq A^b((B(w), w)$ for all $w \in W$ and $A^b(B, w) > A^b((B(w), w)$ at some $w \in W$.*

Informally, an accuracy frame is self-effacing if you can reason from the outset that there's some belief set which is guaranteed to be at least as good as, and possibly better than, wherever the frame takes you. It is easy to see that the accuracy frame described in our proof of proposition 85 is self-effacing, for this reason. In an intuitive sense, it recommends something else over and above itself; following the advice of the frame will make you worse off than some B from the point of view of accuracy. In the remainder of this section, I will *only* consider frame $\cap$-coherent accuracy frames. Just as we motivated $\cap$-coherence by appeal to an analogy with immodesty or strict propriety, we can do likewise with these self-effacing frames. If you care about having an accurate belief set, these frames are — to repeat an earlier example, from D. Lewis (1971) — like a consumer affairs magazine that recommends the opinions of other such magazines as superior. We should not want to follow its advice when buying a new washing machine.

A self-effacing frame represents a bad situation for an agent to be in, insofar as they are concerned to be accurate. Frames where agents are guaranteed to be

a commissive anti-expert represent a bad situation for an agent to be in, full stop. Eventually, we will look more closely at the connections between self-effacing frames and commissive anti-expertise. Before we can do so, we need to spend some more time investigating the properties of self-effacing frames.

2.4.1 Some properties of self-effacing frames

First, we'll note something about the structure of self-effacing frames.

Proposition 93. *There is some frame $\cap$-coherent and self-effacing accuracy frame $\langle W, R, T, F\rangle$ where B outscores all $B(w)$ and $\cap B = W$.*

Proof. The accuracy frame described in the proof of proposition 85 shows this. □

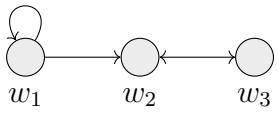
But in general, the belief set B in virtue of which the frame is self-effacing need not be un-opinionated. To see this we first need the following lemma:

Lemma 94. *For all pre-frames $\langle W, T, F\rangle$, if $|F| > (|W| - 1)T$, then if B is consistent and multi-premise closed with respect to W, then B is $\cap$-coherent with respect to itself.*

Proof. Suppose $|F| > (|W| - 1)T$. Then $\frac{|F|}{|F|+T} > \frac{(|W|-1)T}{||W|T} = \frac{|W|-1}{|W|}$. Then by proposition 64, all B that are consistent and multi-premise closed with respect to W are $\cap$-coherent with respect to themselves. □

Proposition 95. *It's not the case for all frame $\cap$-coherent and self-effacing accuracy frames $\langle W, R, T, F\rangle$ where some B scores as well as $B(w)$ at all w, and strictly better than $B(w)$ at some w, that $\cap B = W$.*

Proof. Suppose $|F| > (|W| - 1)T$. Then by lemma 94, for all $B(w)$ in the frame, $B(w)$ is $\cap$-coherent and therefore the illustrated accuracy frame is frame $\cap$-coherent.



$A^b(B(w_1), w_1) = 2T$; $A^b(B(w_2), w_2) = A^b(B(w_3), w_3) = 2T + 2F$. Let $B = \{W, \{w_1, w_2\}, \{w_1, w_3\}, \{w_1\}\}$. Then $A^b(B, w_1) = 4T, A^b(B, w_2) = A^b(B, w_3) = 2T + 2F$. So B is guaranteed to do no worse than any $B(w)$ and may do better: $\langle W, R, T, F\rangle$ is self-effacing. Moreover, $\cap B \subset W$, so B is not un-opinionated.

□

If a frame $\cap$-coherent frame is reflexive, then it is not self-effacing. To show this we first need the following lemma.

Lemma 96. *For all pre-frames, if B is $\cap$-coherent with respect to itself, and $B' = \{p : Pr(p) \geq \frac{|F|}{|F|+T}\}$, where Pr is regular and uniform, then $A^b(B, w) \geq A^b(B', w)$ at all $w \in \cap B$.*

Proof. Since $B' = \{p : Pr(p) \geq \frac{|F|}{|F|+T}\}$, and Pr is regular and uniform, then by lemma 61, the following holds: $(*)$ $A^b(B', w) = A^b(B', w')$ for all $w, w' \in W$.

Now suppose for contradiction that at some $w \in \cap B$, $A^b(B', w) > A^b(B, w)$. Since B is $\cap$-coherent with respect to itself, it is consistent, and therefore scores equally at all $w \in \cap B$. Therefore by $(*)$, $A^b(B', w) > A^b(B, w)$ at all $w \in \cap B$. But then B is strongly $\cap$-dominated with respect to itself, and not $\cap$-coherent. This is a contradiction, so we reject our supposition. □

Proposition 97. *If $\langle W, R, T, F\rangle$ is frame $\cap$-coherent and reflexive, then $\langle W, R, T, F\rangle$ is not self-effacing.*

Proof. Let $B(w)$ be the lowest scoring of all $B(w')$, i.e. $(*)$ for all w', $A^b(B(w'), w') \geq A^b(B(w), w)$. Let $B = \{p : Pr(p) \geq \frac{|F|}{|F|+T}\}$, where Pr is regular and uniform.

Since $\langle W, R, T, F\rangle$ is frame $\cap$-coherent and reflexive, and therefore $w \in \cap B(w)$, by lemma 96, $A^b(B(w), w) \geq A^b(B, w)$. By lemma 61, B scores identically at all $w \in W$, so by $(*)$, for all $w' \in W$, $A^b(B(w'), w') \geq A^b(B, w')$.

By proposition 60, B maximises expected accuracy with respect to Pr. By proposition 55, since B maximises expected accuracy with respect to Pr there is no B' such that $A^b(B', w) \geq A^b(B, w)$ at all w and $A^b(B', w) > A^b(B, w)$ at some w. Since $A^b(B(w), w) \geq A^b(B, w)$ at all $w \in W$, there is no B' such that $A^b(B', w) \geq A^b(B(w), w)$ at all $w \in W$ and $A^b(B', w) > A^b(B(w), w)$ at some $w \in W$. Therefore $\langle W, R, T, F\rangle$ is not self-effacing. □

So every world in a frame $\cap$-coherent accuracy frame accessing itself is sufficient for it to not be self-effacing. We now want to show that it is necessary for the frame $\cap$-coherent frame to not be self-effacing that at least one world sees itself. In more concise terminology:

Definition 98 (Irreflexivity)**.** *A relation R is irreflexive iff for all w, $\langle w, w\rangle \notin R$.*

Definition 99 (Non-irreflexive frame)**.** *An accuracy frame $\langle W, R, T, F\rangle$ is non-irreflexive iff R is not irreflexive.*

We want to show that frame $\cap$-coherent frames must be non-irreflexive if they are not to be self-effacing; equivalently, if they are irreflexive, they are self-effacing. First, we need some additional lemmas.

Lemma 100. *For all pre-frames $\langle W, T, F\rangle$, if B is consistent and multi-premise closed with respect to W, and $w' \notin \cap B$, then w' is in half of $p \in B$.*

Proof. Let B be consistent and multi-premise closed with respect to W. By definition of multi-premise closure, B is the set of all supersets of $\cap B$. Every proposition p in B is the union of $\cap B$ with some element of $\mathcal{P}(W \setminus \cap B)$. Now take some arbitrary w' s.t. $w' \notin \cap B$. This set $\{p \in \mathcal{P}(W \setminus \cap B) : w' \in p\}$ is the union

of $\{w'\}$ with every element of $\mathcal{P}(W \setminus p \cup \{w'\})$. In general, $|\mathcal{P}(X)| = 2^{|X|}$. So $|\mathcal{P}(W \setminus p)| = 2^{|W \setminus p|}$. And $|\mathcal{P}(W \setminus p \cup \{w'\})| = 2^{|W \setminus p \cup \{w'\}|} = 2^{|W \setminus p|-1}$. So w' is in half of $p \in B$. □

Lemma 101. *If B is consistent and multi-premise closed with respect to W, and $w' \notin \cap B$, then $A^b(B, w') = \frac{|B|(T+F)}{2}$.*

Proof. By lemma 100, w' is in half of $p \in B$. So $A^b(B, w') = \frac{|B|}{2}T + \frac{|B|}{2}F = \frac{|B|(T+F)}{2}$.

□

Proposition 102. *If $\langle W, R, T, F\rangle$ is frame $\cap$-coherent but not self-effacing then it is non-irreflexive.*

Proof. We consider two cases corresponding to two classes of A^b: where $T \leq |F|$ and where $|F| < T$. Since $\langle W, R, T, F\rangle$ is frame $\cap$-coherent, then $B(w)$ is consistent and multi-premise closed with respect to W for every w.

First, suppose no world sees itself and $|F| \geq T$. Then by lemma 101, for any given $B(w)$, $A^b(B(w), w) = \frac{|B(w)|(T+F)}{2}$. Since $|F| \geq T$, $\frac{|B(w)|(T+F)}{2}$ is at most zero. So all $B(w)$ score at most zero. But $A^b(\{W\}, w) = T$ at all w, so $\langle W, R, T, F\rangle$ is self-effacing.

Second, suppose no world sees itself and $|F| < T$. Then, as before, for all $B(w)$, $A^b(B(w), w) = \frac{|B(w)|(T+F)}{2}$. Moreover, since $T > |F|$, we know from proposition 64 that by frame $\cap$ coherence of $\langle W, R, T, F\rangle$, $|\cap B(w)| = 1$ throughout. So $|B(w)| = 2^{|W|-1}$. So for all w, $A^b(B(w), w) = \frac{2^{|W|-1}(T+F)}{2} = 2^{|W|-2}(T + F)$; precisely half the members of $B(w)$ are true at w, and precisely half are false.

We now identify a belief set that is guaranteed to score better at every world in

the frame: $\mathcal{P}(W)$. Since $\mathcal{P}(W)$ contains all propositions, it contains more propositions than any $B(w)$; namely, $2^{|W|}$ propositions rather than $2^{|W|-1}$.

But it is easy to see that for any w half of the propositions in $\mathcal{P}(W)$ are true and half are false. So the accuracy of $A^b(\mathcal{P}(W), w) = 2^{|W|-1}(T+F)$. Since $T+F$ is positive, then $2^{|W|-1}(T+F) > 2^{|W|-2}(T+F)$. So $A^b(\mathcal{P}(W), w) > A^b(B(w), w)$ for each w. So $\langle W, R, T, F\rangle$ is self-effacing.

□

This proof shows that if a frame $\cap$-coherent $\langle W, R, T, F\rangle$ is irreflexive, then $\langle W, R, T, F\rangle$ is self-effacing. This result is enough to say something informative about the relationship between anti-expertise and self-effacing frames.

2.4.2 Connections between self-effacing frames and anti-expertise

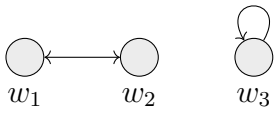
The first thing to notice is that there's a deep connection between self-effacing frames and commissive anti-expertise, to wit:

Proposition 103. *An accuracy frame that represents an agent who is guaranteed to be a commissive anti-expert is a self-effacing frame.*

Proof. That an agent is guaranteed to be a commissive anti-expert about p means $(p \cap \mathbf{B}\neg p) \cup (\neg p \cap \mathbf{B}p)$ is true at every world. So at each world w, either p is true at w and every w' accessed by w is a $\neg p$ world, or $\neg p$ is true at w but every w' accessed by w is a p world. It follows that no world accesses itself. Therefore the frame is irreflexive. Therefore by proposition 102, the frame is self-effacing. □

On the face of it, this is a significant result. If we impose the 'meta' accuracy coherence requirement on the frames described earlier, then guaranteed commissive anti-expertise will be ruled out. An agent can know from the outset that a guarantee of commissive anti-expertise is — from the point of view of being as accurate as possible — a bad deal.

One thing we *can't* conclude from this is that it's bad — from the point of view of accuracy — to think you *may* end up believing that you're a commissive anti-expert with respect to p. The frames we've been looking at, where you're a commissive anti-expert at every world, are ones in which you are guaranteed to become a commissive anti-expert. But there are frames where you might, but are not guaranteed, to become a commissive anti-expert, and the result above says nothing about these. For an example of this kind of frame, we can think back to one we discussed earlier:



At w_1 and w_2 you're a commissive anti-expert with respect to (e.g.) $\{w_1\}$, and moreover, you believe as much. At w_3 you're not a commissive anti-expert with respect to $\{w_1\}$, and you believe as much. In this frame, you're not guaranteed be a commissive anti-expert, and moreover, you're not guaranteed to believe that you're a commissive anti-expert. Yet you *might* end up a commissive anti-expert, and you might end up believing as much.

This is *not* in general a self-effacing frame, even though you may end up believing you are a commissive anti-expert.[8] If we are concerned to try to rule out these frames on grounds of accuracy-conduciveness, this is a bit disappointing. There may, however, be another accuracy-centred reason that the agent involved should avoid such frames. The frame allows that you may or may not end up believing you are a commissive anti-expert. We know moreover that if you believe you are a commissive anti-expert, then you are at either w_1 or w_2. If we restrict our attention to those worlds and the relation that holds between them, we can give another version of the argument we saw before: at w_1 you hold two true beliefs

[8]For instance, when $T > |F|$, see proposition 124 in the appendix.

and two false beliefs, and similarly at w_2. But if you only believe W and $\{w_1, w_2\}$, then you can hold two true beliefs and *no* false beliefs. In other words, you can foresee that if you end up believing you are commissive anti-expert, you'll again be in a self-effacing frame.

This implicitly assumes, of course, a picture of updating; of how the frame changes as more information is gained. However the picture I've effectively assumed is the standard one, from public announcement logic. To introduce it, we need to define a few conditions (Baltag & Renne, 2016, section 2.1).

Definition 104 (Accuracy frame updated on X)**.** *For all* $X \in \mathcal{P}(W)$, $\langle W, R, T, F\rangle[X!] = \langle W[X!], R[X!], T, F\rangle$.

Definition 105 (W updated on X)**.** *For all* $X \in \mathcal{P}(W)$, $W[X!] = \{w : w \in X\}$

Definition 106 (R updated on X)**.** *For all* $X \in \mathcal{P}(W)$, $R[X!]xy$ *iff* Rxy

To give an informal gloss on what these mean: upon receiving the information X, we remove all not-X worlds from the frame, and leave the relations between those remaining worlds intact (Baltag & Renne, 2016, section 2.1).

Now, we can argue more generally that:

Proposition 107. *If* X *is a commissive anti-expertise proposition and* $\langle W, R, T, F\rangle$ *is a frame* $\cap$*-coherent accuracy frame, then* $\langle W, R, T, F\rangle[X!]$ *is a self-effacing accuracy frame.*

Proof. The proof of proposition 103 shows that if an agent is a commissive anti-expert at a world w, then $\neg Rww$. If X is a commissive anti-expert proposition, then at all w in $W[X!]$, $\neg Rww$. Then by proposition 102, $\langle W, R, T, F\rangle[X!]$ is self-effacing; there is some B such that at all $w \in W[X!]$, $(A^b(B, w) \geq A^b(B(w), w)$ and at some $w \in W[X!]$, $A^b(B, w) > A^b(B(w), w)$. □

In sum, while it's fine to be a commissive anti-expert (remember, all that's needed to be a commissive anti-expert is to have a false belief), it's bad from an accuracy-theoretic point of view to be guaranteed to be a commissive anti-expert, and — and more tentatively — to end up believing that you're a commissive anti-expert. In either case, you can know that from the perspective of accuracy you're ending up with a bad deal.

2.4.3 Omissive anti-expertise

There's more to say about anti-expertise, however, since there's another kind of anti-expertise recognised in the literature: omissive anti-expertise. If you're an omissive anti-expert about p, then you can never have a true belief about p: either $\neg p$ and you believe p, or p, but you don't believe p. In other words: $p \leftrightarrow \neg \mathbf{B} p$ (Sorensen, 1988, p. 392).[9] Unlike commissive anti-experts, omissive experts are not guaranteed to believe falsely. But they are guaranteed to not believe truly.

When it comes to our accuracy frames, commissive anti-expertise is (logically) stronger than omissive anti-expertise.

Proposition 108. *If your beliefs are consistent then if you are a commissive anti-expert with respect to p you are an omissive anti-expert with respect to p.*

Proof. If you're a commissive anti-expert with respect to p then $\mathbf{B}p \cap \neg p \cup \mathbf{B}\neg p \cap p$. If your beliefs are consistent then for all q, $\neg(\mathbf{B}q \cap \mathbf{B}\neg q)$. So if p, by commissive anti-expertise, $\mathbf{B}\neg p$, and by consistency, $\neg \mathbf{B} p$. And if $\neg p$, by commissive anti-expertise, then $\mathbf{B}p$. Which is to say, $(p \supset \neg \mathbf{B} p) \cap (\neg p \supset \mathbf{B} p)$; equivalently $(p \supset \neg \mathbf{B} p) \cap (\neg \mathbf{B} p \supset p)$, or $p \leftrightarrow \neg \mathbf{B} p$. Which is to say, you are an omissive anti-expert with respect to p. □

[9] As per fn. 4 earlier, I'm being sloppy here in my use of '$\leftrightarrow$' for the biconditional, and shortly, $\supset$ for the material conditional, but I hope the reader will forgive my sacrifice of precision in the service of readability. I'm also again using the terminology of 'omissive' anti-expertise slightly differently from how Sorensen uses it: see fn. 7.

I'll refer to agents that are omissive anti-experts about p, but not commissive anti-experts about p, as 'mere' omissive anti-experts about p. As with commissive anti-expertise, mere omissive anti-expertise is by itself no great sin: all that's needed is that p is true and you fail to believe it (that secures $p \supset \neg \mathbf{B}p$; for the other direction note that $\neg p$ is false and so $\neg p \supset \mathbf{B}p$ is true). It's a very demanding conception of rationality that punishes ever missing out on a true belief. Guaranteed omissive anti-expertise, however, is more troublesome.

It turns out that there are frames which guarantee you to be an omissive anti-expert about some proposition, but which are nonetheless not self-effacing. So while the link between commissive anti-expertise and self-effacing frames holds, the link to anti-expertise more generally fails.

Proposition 109. *There are frame $\cap$-coherent frames which represent agents that are guaranteed to be omissive anti-experts but are not self-effacing.*

The proof for this requires a few intermediary results, so I defer it to the appendix (section 2.B.2).

This 'meta-reasoning' about the frames means that although considerations of accuracy prohibit guaranteed commissive anti-expertise, they don't prohibit guaranteed omissive anti-expertise.

It's a separate question whether guaranteed commissive anti-expertise means you're in a much worse predicament than guaranteed (mere) omissive anti-expertise. One thing to note is that we may be able to slightly massage the weirdness of the accuracy-theoretic permissibility of these frames that guarantee omissive anti-expertise. First, note again that omissive anti-expertise means you're guaranteed to miss out on true beliefs about p, but you're not guaranteed to have false beliefs about p. Second, we observe that there are no frames representing omissive anti-expertise where $T > |F|$.

Proposition 110. *For all frame coherent accuracy frames $\langle W, R, T, F\rangle$, if $T > |F|$, $\langle W, R, T, F\rangle$ does not represent an agent guaranteed to be a mere omissive anti-expert.*

Proof. Suppose $\langle W, R, T, F\rangle$ guarantees an agent to be a mere omissive anti-expert with respect to p. Then $(p \cap \neg\mathbf{B}p \cap \neg\mathbf{B}\neg p) \cup (\neg p \cap \mathbf{B}p)$ is true at all w. Since $\langle W, R, T, F\rangle$ is frame coherent, then R is serial (by proposition 73) and since $T > |F|$, every world accesses only one world (by proposition 67). By the truth of the omissive anti-expertise proposition at every world and seriality, if there is some $\neg p$ world, then it accesses only a p world. But if there is a p world, w then since $\neg\mathbf{B}p \cap \neg\mathbf{B}\neg p$ is true at w, then w must access both a p world and a $\neg p$ world. But that is prohibited by $T > |F|$ and proposition 67. □

With these observations in place, I suggest that we can tell the following story. Note first that considerations of accuracy only permit an agent to be guaranteed omissive expertise if she is not epistemically risky; viz. $|F| \geq T$. But if she is not epistemically risky, then it is less surprising that she can be okay with omissive expertise, because *that* only guarantees missing out on some true beliefs, and does not guarantee having false beliefs. Part of what it is to dislike being epistemically risky is to prefer the misfortune of missing out on true beliefs over the misfortune of gaining false beliefs. The tolerance of omissive expertise in such cases is perhaps somewhat understandable.

2.5 Moore propositions

You have a 'Moorean' belief if you believe: that p, and that you don't believe that p, or you believe that: $\neg p$, but you believe that p (Moore, 1942).

I think the status of Moorean beliefs for accuracy epistemologists is an inter-

esting question that, to my knowledge, has not received any serious attention.[10] Various philosophers have noted that there's a sense in which you can't believe Moorean propositions like $p \cap \neg \mathbf{B}p$. Exactly how 'can't' should be interpreted is not straightforward, though — the 'can't' could be a matter of psychological necessity, or some further rational prohibition on believing them.[11] One reason for thinking the latter is that Moorean beliefs are guaranteed to be false; if you believe that p then (assuming you never believe both a proposition and its negation) then either your belief that you don't believe that p, or your belief that you believe $\neg p$, will be false (e.g. Sorensen appeals to this; Sorensen, 1988, pp. 37-38). So the Moorean conjunction will be false overall. That is true, so far as it goes, but from an accuracy theorists's perspective that may not be the full story. It may be that in believing p, you also believe something true. Depending on how the value of true and false beliefs are weighed, you might come out with something of value overall. Here's another way of putting things. We saw before that Fitelson and Easwaran (2015) gave an accuracy theoretic treatment of the lottery and preface paradoxes, according to which you could have an accuracy conducive belief set even though at least some of the beliefs in that set were guaranteed to be false (though of course we argued against the theory of accuracy coherence underpinning their explanation). If we think such accuracy-conducive explanations are available, it's worth thinking too about whether there might be accuracy conducive 'justifications' of Moorean beliefs.

We can investigate the relationship between accuracy frames and Moorean beliefs, just as we did with anti-expertise. It's a feature of (the) doxastic logic(s we've been working with) that, so long as the frame is serial, you can never be guaranteed a given Moorean belief, i.e. there's no (serial) frame such that, at every world, a particular Moorean belief is held.

[10] I mentioned this in passing in section 1.5.2.

[11] See Littlejohn (in press, section 1) for good discussion.

Proposition 111. *There is no serial accuracy frame $\langle W, R, T, F\rangle$ such that at each world, $\boldsymbol{B}(\boldsymbol{B}p \cap \neg p)$ is true.*

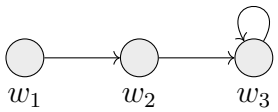
Proof. Suppose $\mathbf{B}(\mathbf{B}p \cap \neg p)$ is true at every world. Then $\mathbf{BB}p \cap \mathbf{B}\neg p$ is true at each world. Since the frame is serial, then by the $\mathbf{BB}p$, every world accessed by each w is a $\mathbf{B}p$ world. Again, since the frame is serial then by $\mathbf{B}\neg p$, every world accessed by each w is a $\neg p$ world. But then no w accesses any p world, and $\mathbf{B}p$ is false at every world. Contradiction. □

Proposition 112. *There is no serial accuracy frame $\langle W, R, T, F\rangle$ such that at each world, $\boldsymbol{B}(p \cap \neg\, \boldsymbol{B}p)$ is true.*

Proof. Suppose $\mathbf{B}(p \cap \neg \mathbf{B}p)$ is true at every world. Then $\mathbf{B}p \cap \mathbf{B}\neg\mathbf{B}p$ is true at every world. Since the frame is serial, then by the truth of $\mathbf{B}p$ at every world, every world only accesses p worlds. Again, since the frame is serial, by the truth of $\mathbf{B}\neg\mathbf{B}p$, every world accesses some world that accesses some $\neg p$ world. Contradiction.

□

There's an unsurprising analogy here with the frames where you were guaranteed to become an commissive anti-expert; if you're guaranteed to have a Moorean belief then the frame will not be accuracy frame $\cap$-coherent, and is therefore bad from the perspective of accuracy. But as with commissive expertise, we can also look at frames where you *might* end up with a Moorean belief, rather than just frames where you are guaranteed to. Here's an example of such a frame; at w_1 you have the Moorean belief that you're at w_2, but that you don't believe you're at w_2. At w_3, you don't have any Moorean beliefs.



As we did with both omissive and commissive anti-expertise, we can ask the dynamic question: what happens, when faced with such a frame, will happen if you end up believing the Moorean proposition?

Proposition 113. *If the Moorean proposition X is true at some $w \in W$, then $\langle W, R, T, F\rangle[X!]$ is not serial.*

Proof. Let X be a Moorean proposition. Then $W[X!]$ is the set of worlds in W at which the Moorean proposition is true, and consequently, a set of worlds at which each world, one holds the Moorean belief in X. But we saw from propositions 111 and 112 when W is as described, there are no such serial frames $\langle W, R, T, F\rangle[X!]$. Consequently, $\langle W, R, T, F\rangle$ is not serial.

□

Since the accuracy frame updated on a Moorean proposition is not serial, it follows by proposition 73 that it is not frame coherent.

I think this is at least some reason to think that agents concerned with accuracy should want to avoid frames that license Moorean beliefs: they know that there's a risk of ending up at a frame that's not frame $\cap$-coherent. However I am quite cautious about drawing too strong a conclusion from this result. It's a general feature of the public announcement logic we've been working with that updating on a serial frame is not guaranteed to return a serial frame (as our example in effect shows), so updating on a $\cap$-coherent frame is not guaranteed to preserve $\cap$-coherence. Once we realise that frame $\cap$-coherence is not preserved under updates, we might question whether it is really so bad that the possibility of Moorean belief can lead you to end up at a non-frame $\cap$-coherent frame. All kinds of updates, including non-Moorean ones, may do so.

I can only gesture at a response. Given a broader picture of $\cap$-coherence, we should want a full dynamic account, in addition to the static account we have pur-

sued. I have only gestured at a dynamic account here; as I will discuss later, a full such account is an important topic for future research. One thing this dynamic account should hope to secure is an account of updating wherein frame $\cap$-coherence is preserved under updates (by way of example, if a frame is serial *and* euclidean, then the updated frame will be, too; so there are at least some ways of preserving seriality (Baltag & Renne, 2016, section 2.1)). However what we showed above is that there's no way to ensure that 'Moorean' updates will lead to serial frames: so even when (if) we tell a full dynamic story, we know that this story is one that Moorean updates will run afoul of. This is considerably more impressionistic than our treatment of anti-expertise. Moreover, it relies upon there being some plausible dynamic story to be found in this setting. I think the most we can say with any real conviction is that there is at least a possible story as to why it is bad to end up holding Moorean beliefs that is grounded in the notion of $\cap$-coherence.

2.6 An alternative picture

Let's recap what we've learned so far. By looking at accuracy frame $\cap$-coherent frames, we find no justification for belief introspection principles, and considerations — of varying strengths — against anti-expertise and Moorean beliefs. Earlier I noted that in principle we might like to find accuracy-centred reasons in favour of the introspection principles, and to have more consistent considerations against anti-expertise and Moorean beliefs. In fact, there is another way of examining these questions that seems to get us at least some of the way there, by apparently vindicating the introspection principles. In this section I outline this framework, and express some reservations about thinking of the problems we've been looking at in this way.

Our discussion in the paper so far has been indirectly decision-theoretic; we

took the $\cap$-coherence normative requirement — grounded in dominance principles — which we then imposed on the frames, before asking questions about introspection principles. But, at least on the face of it, we could ask more directly whether dominance reasoning bears on introspection. There are a few ways we could set up the decisions for the agents in question. Here's the first thing we might try:

	$\mathbf{B}p$	$\neg\mathbf{B}p$
$\mathbf{BB}p$	T	F
$\neg\mathbf{BB}p$	0	0

Suppose you're deciding whether to believe that you believe p. Then whether your higher-order belief would be true depends on whether you in fact believe p. As the table describes, believing that you believe p isn't dominated by not so believing. But equally, refraining from higher-order belief is also not dominated. On the face of it, the higher-order belief is permissible, but not obligatory.

It may be tempting to think that $\neg\ \mathbf{B}p$ should not even be a live proposition. On the approach we're sketching here, I think that temptation should be rejected, for three reasons. First, we're interested in whether there are accuracy-centred reasons for holding higher-order beliefs about your beliefs. If we assume from the outset that your beliefs rule out the possibility of being wrong about your beliefs, we learn nothing.

Second, informal reflection reveals that there are many occasions when we have beliefs but don't realise we have them. I watch a tennis match, about the players of which I am ostensibly neutral. As the match progresses, I find myself disappointed that one of the players is losing. It turns out that I believe *they* deserved to win, though I didn't believe that I held that belief.

Third, there are more rigorous, philosophical reasons for thinking that we don't in general have perfect access to our own mental states. One example is

Timothy Williamson's anti-luminosity arguments which apparently show there are no non-trivial mental states such that we always have perfect access to them (Williamson, 2000, ch. 4). Although Williamson's arguments are couched in terms of *knowing* whether we have certain mental states, other philosophers have suggested that the margins-for-error type principles underpinning Williamson's argument hold for (rational) belief (e.g. Schultheis, 2018). Another is the (quasi)-empirical data against our introspective skills canvassed in Schwitzgebel (2006).

We can give a similar decision-theoretic argument for the permissibility, though not requirement, of higher-order beliefs about what you *don't* believe.

	$\mathbf{B}p$	$\neg\mathbf{B}p$
$\mathbf{B}\neg\,\mathbf{B}p$	F	T
$\neg\mathbf{B}\neg\mathbf{B}p$	0	0

Once more, in both cases the higher-order belief is undominated. So either are permissible. But neither is required.

If we specify the decision problems differently, we can get slightly different results. Instead of thinking about forming the higher-order belief independently of the first-order belief, we can also examine what happens if you form them together.

	p	$\neg p$
$\mathbf{BB}p\cap\mathbf{B}p$	$2T$	$T+F$
$\neg\mathbf{BB}p\cap\mathbf{B}p$	T	F
$\neg\mathbf{BB}p\cap\neg\mathbf{B}p$	0	0

This table tells us two interesting things. First, and in effect as we saw in the previous chapter, whether you should believe p in the first place will be sensitive to the values T and F. If $T < |F|$ then not believing it will be undominated. Second, and more interestingly, though, the table seems to show that it ought to be the case that, if you believe p, you believe that you believe p. Suppose you

believe p, but don't believe that you believe p. Then on the face of it, you will be accuracy dominated; whether you are at a world in p, or in $\neg p$, by believing p without believing that you believe it, you are guaranteed to be missing out on accuracy.

It's worth comparing this to the first decision table we saw. That seemed to show that it's *not the case* that if you believe p, you ought to believe that you believe p. On the other hand, our second argument shows that it ought to be the case that if you believe p, you believe that you believe p. These two different requirements exhibit what is known as narrow versus wide scope. In the former case, the 'ought' scopes over only the consequent, in the latter case it scopes over the whole conditional (Broome, 2007). Interestingly, this parallels the accuracy arguments given for single-premise closure. It's not in general the case that, if you believe p, then you ought to believe q on pain of accuracy dominance. The belief set with q added may not dominate the former, unclosed set (Easwaran, 2016, pp. 825-26). It is in general the case that, it ought to be that if you believe p, then you believe q. Any belief set which fails to satisfy this will be accuracy dominated, as Easwaran has shown. In that sense, we might think that failure of our first attempt to justify a kind of belief introspection requirement is unsurprising is just a broader symptom of the fact it is difficult to get interesting narrow-scope requirements out these kinds of accuracy frameworks.

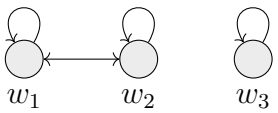
Just as we argued for positive introspection, we can argue for negative introspection.

	p	$\neg p$
$\mathbf{B}\neg\mathbf{B}p \cap \neg\mathbf{B}p$	T	T
$\neg\mathbf{B}\neg\mathbf{B}p \cap \neg\mathbf{B}p$	0	0

This has the same consequence; it ought to be that if you fail to believe p, you believe that you fail to believe it. Once more, the wide-scope version of this

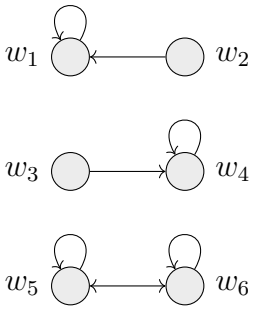
constraint holds, while the narrow scope doesn't.

Despite these apparently impressive results, I have some reservations about looking at things this way. In brief, I am concerned that these decision tables make tacit assumptions about the background doxastic logic, and once those assumptions are made explicit, the various arguments no longer go through. Here's an example of what I mean. In the following frame, you believe $\{w_1, w_2\}$ if and only if you believe that you believe $\{w_1, w_2\}$.



So in fact, the act corresponding to the row in the decision table where you believe the proposition but do not believe that you believe it is simply not possible. The decision problem is not well-posed. This by itself isn't a decisive problem for this way of looking at things — not least because in this example the accessibility relation is an equivalence relation — but it does highlight that at least some care will be needed in thinking about the decision problems.

If we do take some more care, and try to ensure the decision problems of interest include all relevant possible acts, it's not clear we get the desired results after all. Here's an example of how we could handle this, which doesn't yet allow for believing p but not believing that you believe it.



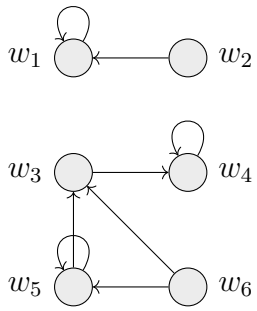
Let $p = \{w_1, w_3, w_5\}$, i.e. the left column of worlds, and $\neg p = \{w_2, w_4, w_6\}$ i.e. the right column of worlds. Believing p corresponds to the first row, believing $\neg p$ to the second, and believing neither to the third.

From lemma 100 earlier, we know that the values in the decision table will be as in table 2.1.

	p	$\neg p$
$\mathbf{B}p$	2^5T	$2^4(T+F)$
$\mathbf{B}\neg p$	$2^4(T+F)$	2^5T
$\neg\mathbf{B}p \cap \neg\mathbf{B}\neg p$	2^4T	2^4T

Table 2.1: Revised decision table 1

Each belief set is undominated. Now the general picture is on the table, we can look more carefully at decision problems where believing, but not believing that you believe, are available.



Let the left and right column stand for p and $\neg p$, as before. Then the first row amounts to believing p, and believing that you believe it, and the second to believing $\neg p$, and believing that you believe it. The final row amounts to believing p — both worlds only see worlds in the left column — but not believing that you believe it — both worlds see w_3, where you do not believe p. Again, given lemma 100, we know that the resulting decision table will be as in table 2.6.

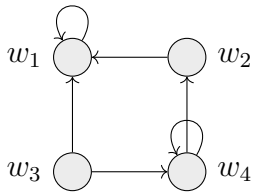
Notice that the first row — believing and believing that you believe — will weakly dominate the third row — believing without believing that you believe

	p	$\neg p$
$\mathbf{B}p \cap \mathbf{BB}p$	2^5T	$2^4(T+F)$
$\mathbf{B}\neg p$	$2^4(T+F)$	2^5T
$\mathbf{B}p \cap \neg\mathbf{BB}p$	2^4T	$2^3(T+F)$

Table 2.2: Revised decision table 2

— only if $T > F$. Otherwise, all actions are undominated. This is enough to show that in general, thinking about the positive introspection principle in this alternative way is not, by itself, sufficient to give an accuracy-theoretic vindication of it.

There's a similar example for the negative introspection principle.



Again letting the columns correspond to p and $\neg p$ respectively, the first row amounts to not believing $\neg p$, but believing that you don't believe $\neg p$; the second amounts to not believing $\neg p$, but also, not believing that you don't believe $\neg p$. Yet:

	p	$\neg p$
$\mathbf{B}p \cap \mathbf{BB}p$	2^3T	$2^2(T+F)$
$\mathbf{B}\neg p$	$2(T+F)$	2^2T

Table 2.3: Revised decision table 3

So neither is dominated.

Given these results, my tentative diagnosis of the earlier arguments would be that they're insufficiently careful about pulling apart the relevant believed (or unbelieved) propositions. Once we are more careful, the arguments don't go through. I emphasise 'tentative'. It would not completely surprise me if there were other

ways of understanding some of the problems in this section which yielded different results. More generally, I think how the two different approaches we've discussed are related is an important question for future research. My modest conclusion at this point would just be that the frame-based approach which has taken up the majority of this chapter is a) not clearly mistaken and b) of independent interest regardless. I say this while keeping a very open mind about the prospects of there being other ways of pursuing the questions we've looked at, with interestingly different outcomes.

2.7 Conclusion

We've seen that $\cap$-coherence, when integrated with doxastic logic frames, can give us the resources to ask (and answer) a host of interesting questions about higher-order belief. It turns out that, so formalised, there are no accuracy-theoretic reasons to hold so-called 'belief introspection' principles. There are accuracy-theoretic reasons to avoid thinking of yourself as a commissive anti-expert, and though there aren't in general reasons to avoid thinking of yourself as an omissive anti-expert, we can make some sense of this by paying attention to how epistemically risky (or not) you are. Lastly, we saw that there are (admittedly impressionistic) accuracy-centred reasons to avoid holding Moorean beliefs.

This appendix collects various — mostly technical — expansions of the issues discussed above.

Appendix 2.A Accuracy frame W-coherence

In the previous chapter we discussed an alternative picture of accuracy coherence — W-coherence — discussed most prominently in Easwaran (2016) and Fitelson

and Easwaran (2015). We now discuss this notion in more detail, and apply it to accuracy frames.

Definition 2 (Strong W-dominance)**.** *B' strongly W-dominates B iff $A^b(B', w) > A^b(B, w)$ at all $w \in W$.*

Definition 3 (Weak W-dominance)**.** *B' weakly W-dominates B iff $A^b(B', w) \geq A^b(B, w)$ at all $w \in W$ and $A^b(B', w) > A^b(B, w)$ at some $w \in W$.*

These are coupled with the corresponding normative claim:

> **W-coherence:** It's rationally prohibited for W to be either strongly or weakly W-dominated with respect to W.

Whereas $\cap$-coherence requires you to do as well as possible at those worlds left open to you, W-coherence requires you to do as well as possible with respect to every possible world. Proposition 55 therefore shows that if B maximises expected accuracy relative to some regular Pr, B is neither strongly nor weakly W-dominated: B is W-coherent. As we mentioned before, it is because Easwaran is only interested in W-dominance that he doesn't consider irregular probability functions.

We can elucidate the relationship between the two notions of accuracy coherence.

Proposition 114. *If B is $\cap$-coherent with respect to itself then there is no B' that strongly W-dominates B.*

Proof. Since B is $\cap$-coherent with respect to itself, there is no B' such that $A^b(B', w) > A^b(B, w)$ at all $w \in \cap B$. Since $\cap B \subseteq W$, it follows that there is no B' such that $A^b(B', w) > A^b(B, w)$ at all $w \in W$; i.e. there is no B' that strongly W-dominates B. □

Interestingly, this doesn't extend to weak W-dominance:

Proposition 115. *There are belief sets that are $\cap$-coherent with respect to themselves but weakly W-dominated.*

Proof. Let $W = \{w_1, w_2, w_3\}$, $B = \{W, \{w_1, w_2\}\}$, and $T = |F|$. We first show B is $\cap$-coherent with respect to itself. Let $Pr(\{w_1\}) = Pr(\{w_2\}) = \frac{1}{2}$. It is easy to see that for all $p \in B$, $Pr(p) \geq \frac{|F|}{|F|+T} = \frac{1}{2}$, and for no $p \notin B$, $Pr(p) > \frac{1}{2}$. Therefore by Proposition 60, B maximises expected accuracy relative to Pr, and can be neither weakly nor strongly $\cap$-dominated with respect to itself.

We now show that B is weakly W-dominated. Let $B' = \{W, \{w_1, w_2\}, \{w_1, w_3\}, \{w_2, w_3\}\}$. Then $A^b(B', w_1) = A^b(B', w_2) = A^b(B', w_3) = 3T + F = 2T$. Whereas $A^b(B, w_1) = A^b(B, w_2) = 2T$, and $A^b(B, w_3) = T + F$, which is strictly less than $2T$. So B' weakly W-dominates B. □

We now define a related notion of frame coherence:

Definition 116. *A frame $\langle W, R, T, F\rangle$ is frame W-coherent iff for all w, $B(w)$ is W-coherent.*

That $\cap$-coherence of a belief set with respect to itself doesn't entail W-coherence of a belief set means we cannot straightforwardly infer that a frame $\cap$-coherent frame is also frame W-coherent. However we can make more restricted inferences. Recall proposition 75 above: If B is consistent, multi-premise closed with respect to W and $|\cap B| = 1$ then for all T, F, B is $\cap$-coherent with respect to itself.

We add to this the following observation:

Proposition 117. *If B is consistent, multi-premise closed with respect to W and $|\cap B| = 1$ then for all T, F, B is W-coherent.*

Proof. Suppose B is consistent, multi-premise closed with respect to W and $|\cap B| = 1$. Then B contains all and only propositions true at the $w \in \cap B$. Any proper subset $X \subset B$ scores worse at than B at w. Any subset $Y \subseteq W \setminus B$ contains only propositions false at w, so scores worse at w. And therefore any $X \cup Y$ scores worse than B at w. But that exhausts all possible belief sets. Therefore no B' scores even at least as well as B at w, and B is neither strongly nor weakly W-dominated. □

From this we can show:

Proposition 118. *For all accuracy frames where, for all w, $|B(w)| = 1$, that frame is frame $\cap$-coherent and frame W-coherent.*

Proof. Follows straightforwardly from propositions 75 and 117. □

From this we can show that shifting to frame W-coherence makes little difference to the questions we asked about belief introspection.

Proposition 119. *There are no values of T and F such that, all frame W-coherent accuracy frames with those values are transitive.*

Proof. Follows from proposition 85. □

Proposition 120. *There are no values of T and F such that, all frame W-coherent accuracy frames with those values are euclidean.*

Proof. Follows from proposition 86. □

So altering the notion of belief coherence from $\cap$-coherence to W-coherence makes no difference to whether one should positively or negatively introspect.

Since, as proposition 115 shows, frame $\cap$-coherence does not entail frame W-coherence, this result is somewhat interesting.[12]

Similarly, assuming W-coherence and either transitivity or the euclidean property does not let us infer anything about the values of T and F.

Proposition 121. *There exists an accuracy frame $\langle W, R, T, F\rangle$ such that $\langle W, R, T, F\rangle$ is transitive, W-coherent and for all T', F', $\langle W, R, T', F'\rangle$ is W-coherent.*

Proof. As above. □

Proposition 122. *There exists an accuracy frame $\langle W, R, T, F\rangle$ such that $\langle W, R, T, F\rangle$ is euclidean, W-coherent and for all T', F', $\langle W, R, T', F'\rangle$ is W-coherent.*

Proof. As above. □

Let's take stock. We can shift from imposing the requirements of $\cap$-coherence on doxastic logic frames to W-coherence. But as before, that logic is weak, requiring neither transitivity nor the eucldiean property. On this way of looking at things, there are no accuracy-motivated belief introspection principles.

Proposition 115 has another interesting consequence. We know from proposition 63 that whenever a belief set B maximises expected accuracy with respect to Pr_B, it is $\cap$-coherent with respect to itself. It might be tempting to think that in these cases, we could siphon off some $\epsilon > 0$ probability from $\cap B$ and distribute it over all the worlds in $W \setminus \cap B$, in such a way that B also maximises expected

[12]Say that a belief set B is represented in an accuracy frame iff for some w, $B(w) = B$. One salient difference between frame $\cap$-coherence and W-coherence is that there are no belief sets that are $\cap$-coherent with respect to themselves that cannot be represented in some accuracy frame. However there are W-coherent belief sets that cannot be represented in some accuracy frame. This is because, as we have seen, W-coherence permits failures of multi-premise closure. Our accuracy frames impose this constraint on all belief sets. This difference can be ignored, however, since we can, as we've seen, represent *some* W-coherent belief sets in accuracy frames, and there are *some* frame W-coherent accuracy frames. We can use these frames to provide the relevant counterexamples to certain universal claims about frame W-coherence, as in the case of propositions 119 and 120.

accuracy with respect to this regular Pr'. But proposition 115 suffices to show this is not, in general, possible. There are B which are $\cap$-coherent with respect to themselves, and therefore maximise expected accuracy with respect to Pr_B, but which do not maximise expected accuracy relative to any Pr which does not assign probability one to $\cap B$.

Proposition 123. *It is not the case that for all B that are $\cap$-coherent with respect to themselves, there is some regular Pr such that B maximises expected accuracy with respect to Pr*

Proof. Here is an example. Let $T = |F|$, and $W = \{w_1, w_2, w_3\}$. $B = \{W, \{w_1, w_2\}\}$ maximises expected accuracy with respect to the probability function Pr such that $Pr(\{w_1\}) = Pr(\{w_2\}) = \frac{1}{2} = \frac{|F|}{|F|+T}$, since by proposition 60, for all $p \in B$, $Pr(p) \geq \frac{|F|}{|F|+T}$ and for no $p \notin B$, $Pr(p) > \frac{|F|}{|F|+T}$.

Now suppose $Pr'(\{w_1, w_2\}) < 1$. Then either $Pr'(\{w_1\}) < \frac{1}{2}$ or $Pr'(\{w_2\}) < \frac{1}{2}$. Therefore either $Pr'(\{w_2, w_3\}) > \frac{1}{2}$ or $Pr'(\{w_1, w_3\}) > \frac{1}{2}$. So there is some $p \notin B$ such that $Pr'(p) > \frac{1}{2}$. Therefore by proposition 60, B cannot maximise expected accuracy with respect to any probability function Pr' such that $Pr'(\{w_1, w_2\}) < 1$. □

Appendix 2.B More results about self-effacing frames

We investigated the self-effacing frames above when thinking about commissive anti-expertise. But the frames are interesting in their own right. Here, we prove some further results about the frames.

2.B.1 Sufficient conditions for avoiding self-effacing frames when $T > |F|$

We have already above seen that it is sufficient for a frame $\cap$-coherent accuracy frame to not be self-effacing that it is reflexive. A natural hypothesis is that a weaker sufficient condition is that a certain number of worlds must see themselves. This is the case whenever $T > |F|$.

Proposition 124. *If $T > |F|$ and $\langle W, R, T, F\rangle$ is frame $\cap$-coherent, then if $\langle W, R, T, F\rangle$ is self-effacing, then $\langle W, R, T, F\rangle$ is irreflexive.*

Proof. Suppose that $T > |F|$, $\langle W, R, T, F\rangle$ is frame $\cap$-coherent and non-irreflexive. Since $T > |F|$, by proposition 67, each world accesses exactly one world. By non-irreflexivity of the accuracy frame, some worlds access themselves, and since $T > |F|$, only access themselves. Let w be such a world.

There are now two cases to consider: either $\langle W, R, T, F\rangle$ is self-effacing in virtue of some B, where $B \neq B(w)$ or it is self-effacing in virtue of some B, where $B = B(w)$.

Since $|\cap B(w)| = 1$, and $w \in \cap B(w)$, it follows from proposition 117 that there is no $B \neq B(w)$ such that $A^b(B, w) \geq A^b(B(w), w)$. Moreover, when $B = B(w)$, $A^b(B, w) = A^b(B(w), w)$, but $A^b(B, w') \leq A^b(B(w'), w')$ for all $w' \in W \setminus \{w\}$. Therefore in either case, it is not the case that $A^b(B, w) \geq A^b(B(w), w)$ at all $w \in W$, and $A^b(B, w) > A^b(B(w), w)$ at some $w \in W$. Therefore $\langle W, R, T, F\rangle$ is not self-effacing. □

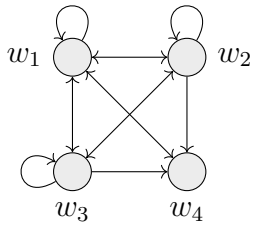
So some world accessing itself is necessary and sufficient for a frame $\cap$-coherent frame to avoid being self-effacing, whenever $T > |F|$.

Interestingly, though, this is not always the case when $|F| \geq T$. At least relative to some such values, there can be (finitely) arbitrarily many worlds short of all that access themselves, and the frame can still be self-effacing.

Proposition 125. *For any finite n, there is some $\langle W, R, T, F\rangle$ such that for $n-1$ of the $w \in W$, Rww, $\langle W, R, T, F\rangle$ is frame $\cap$-coherent and $\langle W, R, T, F\rangle$ is self-effacing.*

Proof. Let $|F| > (|W|-1)T$. Then by lemma 94, any belief set that is consistent and multi-premise closed with respect to W is $\cap$-coherent with respect to itself. Therefore relative to these values, any serial frame is frame $\cap$-coherent. Let $|W|-1$ of the worlds w access every world. Let the remaining world w' access only one distinct world, i.e. one of the $|W|-1$ worlds. Since all of the first $|W-1|$ worlds access every other world, then $B(w) = W$, and $A^b(B(w), w) = T$ for those worlds. Now consider w'. w' only accesses some distinct world. So $A^b(B(w'), w') = \frac{|B|(T+F)}{2}$. Since $|F| > T$, $B(w')$ receives a negative score at w'. So $\{W\}$ does at least as well as $B(w)$ at every world in the frame, and better at one – namely, better at w'. So $\langle W, R, T, F\rangle$ is self-effacing. □

Here is an example of the above using an accuracy frame where $n = 4$.



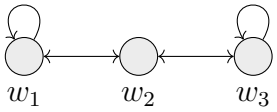
So whatever the full story is about which conditions are needed to prevent a frame $\cap$-coherent accuracy frame from being self-effacing, it is not just a matter of the frame being 'accurate enough', in the sense of enough worlds seeing themselves. At least when $|F|$ is at least as great as T, the story must be more complicated.

2.B.2 Sufficient conditions for avoiding self-effacing frames when $|F| \geq T$

In fact, things are considerably more complicated. First I describe in some detail one sufficient condition. Then I show it is a special case of a much more general, though less informative, condition, which is both necessary and sufficient.

To do this we need to introduce some new properties.

Definition 126 (Matching)**.** *A frame* $\langle W, R, T, F\rangle$ matches *a belief set* B *with respect to* $X \subseteq W$ *iff* $A^b(B(w), w) = A^b(B, w)$ *at all* $w \in X$.

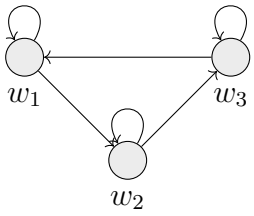


This first frame *matches* (at least) $B = \{W, \{w_1, w_3\}\}$ with respect to W. For $A^b(B(w_1), w_1) = A^b(B(w_3), w_3) = A^b(B, w_1) = A^b(B, w_3) = 2T$. And $A^b(B(w_2), w_2) = A^b(B, w_2) = T + F$. Note that the frame matches B even though for all w, $B(w) \neq B$; a matched belief set need not be represented in the frame.

Here is a related property.

Definition 127 (Superseding)**.** *A frame* $\langle W, R, T, F\rangle$ supersedes *a belief set* B *with respect to* $X \subseteq W$ *iff* $A^b(B(w), w) \geq A^b(B, w)$ *at all* $w \in X$ *and* $A^b(B(w), w) > A^b(B, w)$ *at some* $w \in X$.

In this second frame, $A^b(B(w_1), w_1) = A^b(B(w_2), w_2) = A^b(B(w_3), w_3) = 2T$. Therefore the frame *supersedes* $B = \{W, \{w_1, w_3\}\}$ with respect to W, for as we saw above, $A^b(B, w_2) = T + F$.



We can use these properties to shed some light on the self-effacing frames we have been discussing.

Proposition 128. *If $\langle W, R, T, F\rangle$ is frame $\cap$-coherent, B is $\cap$-coherent with respect to itself, and $\langle W, R, T, F\rangle$* supersedes *$B$ with respect to $\cap B$, then $\langle W, R, T, F\rangle$ is not self-effacing.*

Proof. Since $\langle W, R, T, F\rangle$ supersedes B with respect to $\cap B$, $A^b(B(w), w)$ is at least as great as $A^b(B, w)$ for all $w \in \cap B$, and strictly greater at some such world. B is $\cap$-coherent with respect to itself, so by proposition 49 there's no belief set that strongly W-dominates B. Proposition 50 allows that B may be weakly W-dominated, only by some rival belief set outscoring B at some world not in $\cap B$. Since B is $\cap$-coherent with respect to itself, any rival belief set at best scores equally at those $w \in \cap B$. Since $\langle W, R, T, F\rangle$ supersedes B, $A^b(B(w), w) > A^b(B, w)$ at some $w \in \cap B$. So there can be no belief set B' that scores at least as well as $B(w)$ at all worlds in $\cap B$ and better than some. So $\langle W, R, T, F\rangle$ is not self-effacing. □

What this result does is find some relationship between a frame and a belief set that is $\cap$-coherent with respect to itself. Because that belief set is $\cap$-coherent with respect to itself, and because the frame supersedes that set with respect to $\cap B$, no rival belief set does at least as well at all worlds and better at some. Because of this, we needn't look for belief sets that are $\cap$-coherent with respect to themselves. We can look directly for frames that match or supersede W-coherent belief sets, at all $w \in W$. In other words:

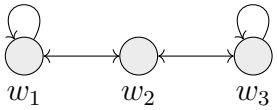
Proposition 129. *$\langle W, R, T, F\rangle$ is frame $\cap$-coherent and not self-effacing iff $\langle W, R, T, F\rangle$* matches *or* supersedes *some B that is W-coherent, relative to W.*

Proof. As described above. □

This allows us to finally prove proposition 109.

Proposition 109. *There are frame $\cap$-coherent frames which represent agents that are guaranteed to be omissive anti-experts but are not self-effacing.*

Proof. Observe the following frame.



First we verify that the agent is guaranteed to be an omissive anti-expert with respect to the proposition $\{w_1, w_3\}$. At w_1 and w_3, the proposition is true but the agent doesn't believe it. At w_2, the proposition is false, but the agent believes it. Therefore at every world, $\{w_1, w_3\} \leftrightarrow \neg\mathbf{B}\{w_1, w_3\}$ is true, as required.

Next, we show that the frame is frame coherent. Suppose $T = 1$ and $F = -5$. Then $\frac{|F|}{|F|+T} = \frac{5}{6}$. Now suppose $Pr(\{w_1\}) = Pr(\{w_2\}) = \frac{1}{2}$. Therefore $Pr(p) \geq \frac{5}{6}$ iff $p = \{w_1, w_2\}$ or $p = W$. Therefore by proposition 63, the belief set $B(w_1) = \{W, \{w_1, w_2\}\}$ is $\cap$-coherent with respect to itself. Parallel arguments show that the belief sets $B(w_2)$ and $B(w_3)$ are $\cap$-coherent with respect to themselves. Therefore the frame is frame coherent.

Lastly, we show that the frame is not self-effacing. Suppose $Pr(\{w_1\}) = Pr(\{w_2\}) = \frac{1}{6}$. Therefore $Pr(W) \geq Pr(\{1, 3\}) \geq \frac{5}{6}$, and all other propositions p are such that $Pr(p) < \frac{5}{6}$. Therefore $\{W, \{w_1, w_3\}\}$ is W-coherent. Therefore by proposition 129, the frame is not self-effacing. □

Proposition 129 gives a schematic explanation of whether a frame is self-effacing. To have a more detailed explanation, we need to have both necessary and sufficient conditions for W-coherence of a given B.

Some necessary conditions are already known. For example, and as we saw in the previous chapter, Easwaran (2016, p. 825) proves that if B is W-coherent then B is single-premise closed. Since accuracy frames impose multi-premise closure on the belief sets they represent, this condition is automatically satisfied. We also know some sufficient conditions. We have seen already that if B is consistent, multi-premise closed with respect to W and $|\cap B| = 1$, then B is W-coherent. Similarly, if B maximises expected accuracy with respect to a regular probability function, then per lemma 54, B is W-coherent. But neither of these conditions are necessary. Easwaran (2016, p. 824) shows (in effect) that, at least for some pre-frames, and in particular some values of T and F, a) there are W-coherent belief sets that are not consistent, and therefore not $\cap$-coherent with respect to themselves, and b) there are W-coherent belief sets that do not maximise expected accuracy relative to any probability function.

My suspicion is that it will be very difficult to give informative joint necessary and sufficient conditions for W-coherence. The reason is that equivalent problems remain, as far as I can tell, open research-mathematical questions. Let me sketch this point in some more detail.

We first need to re-describe some of our framework in a manner more amenable to the credal accuracy literature. In that literature, the accuracy of credence functions $Cr\colon\ \mathcal{P}(W) \rightarrow [0,1]$ are evaluated using credal inaccuracy measures A^c, which measure the distance of a given credence function from the 'omniscient' credence function i.e. the truth function at a world — more on this later.[13] We can

[13]Note 'distance' may be misleading since these measures typically don't satisfy the triangle inequality. The measures are typically Bregman divergences. See Pettigrew (2016a) for more on this.

set up our accuracy framework for outright belief in a similar way.

For notational ease, let $n = |\mathcal{P}(W)|$, i.e. n is the size of the algebra our belief sets are defined over. We then correspond to each proposition p in $\mathcal{P}(W)$ a vector $\mathbf{p}$ in the space $\{0,1\}^n$. Each propositional vector takes the co-ordinate 1 for the dimension corresponding to that proposition and 0 otherwise.

We correspond to each belief set B an n-dimensional vector $\mathbf{b} := \langle p_1, p_2, ..., p_n \rangle$ where $p_i = 1$ iff $p_i \in B$ and 0 otherwise.

We now need a way of measuring the distance between belief sets. First, we measure the distance between individual beliefs using a scoring rule, s^b.

Definition 130 (Scoring rule)**.**

$$s^b(\mathbf{b}, \mathbf{b}', p_i) = \begin{cases} T & \textit{ith co-ordinate in } \mathbf{b} - \textit{ith co-ordinate in } \mathbf{b}' = 1 \\ |F| & \textit{ith co-ordinate in } \mathbf{b} - \textit{ith co-ordinate in } \mathbf{b}' = -1 \\ 0 \textit{ otherwise} & \end{cases}$$

The distance D between two belief sets is then the sum of the distances between all of their constituent beliefs:

Definition 131.

$$D(\mathbf{b}, \mathbf{b}') := \sum_{p \in \mathcal{P}(W)} s(\mathbf{b}, \mathbf{b}', p)$$

Lastly, we identify the omniscient belief sets: all and only the maximally consistent belief sets. We label these $\mathbf{b}_{\mathbf{w}_1}$, $\mathbf{b}_{\mathbf{w}_2}$ etc. corresponding to the strongest proposition they believe.

The inaccuracy of a belief set at a world is now found by taking the distance between the omniscient belief set at that world and the belief set we are interested in. Here is a quick toy example. Let $W = \{w_1, w_2\}$, $B = \{W\}$. We can measure the distance to B from the omniscient belief set at w_1 as follows:

$$D(\mathbf{b}_{\mathbf{w}_1}, \mathbf{b}) = \sum_{p \in \mathcal{P}(W)} s(\mathbf{b}_{\mathbf{w}_1}, \mathbf{b}, p)$$

Ordering the propositions in $\mathcal{P}(W)$ in the obvious way: $\langle \emptyset, \{w_1\}, \{w_2\}, W \rangle$, then:

$\mathbf{b}_{\mathbf{w}_1} = \langle 0, 1, 0, 1 \rangle$ and $\mathbf{b} = \langle 0, 0, 0, 1 \rangle$.

So: $s^b(\mathbf{b}_{\mathbf{w}_1}, \mathbf{b}, \emptyset) = 0$, $s^b(\mathbf{b}_{\mathbf{w}_1}, \mathbf{b}, \{w_1\}) = T$, $s^b(\mathbf{b}_{\mathbf{w}_1}, \mathbf{b}, \{w_2\}) = 0$, $s^b(\mathbf{b}_{\mathbf{w}_1}, \mathbf{b}, W) = 0$. Summing, we get:

$D(\mathbf{b}_{\mathbf{w}_1}, \mathbf{b}) = T$

This gives us an *inaccuracy* score — it tells us how far short the belief set $\mathbf{b}$ falls from perfect accuracy at w_1. We can recover the accuracy scores we've been working with so far by subtracting the inaccuracy score from the greatest possible accuracy score at that world, e.g. $2^{(n-1)}T$. In our case, since $n = 2$, $2^1 T - T = T$.

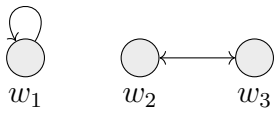
The task of finding W-coherent belief sets then, amounts to finding the pareto set; the set of belief sets such that, for every belief set in that set, no belief set is closer to all omniscient belief sets. This is non-trivial, and closely related to problems in operations research such as the Fermat-Weber (or facility location) problem. I suspect giving a very general account of W-coherence will prove difficult.

Appendix 2.C Self-effacing frames and belief introspection

We saw in our discussion of anti-expertise that we can use avoidance of self-effacing frames as a sort of accuracy consideration at one remove, to justify further principles (like not believing you're a commissive anti-expert). We should check whether the 'meta' requirement motivates belief introspection principles.

Proposition 132. *There are no values of T and F such that, all frame $\cap$-coherent and non-self-effacing accuracy frames with those values are transitive.*

Proof. We describe the following frame.



This accuracy frame is *robust* in the sense of definition 74 since for all w, $|\cap B(w)| = 1$. It is also not self-effacing. For the case where $T > F$, we have shown that it is sufficient for a frame to not be self-effacing that one world accesses itself. For the case where $|F| < T$, it is easy to see that $B(w_1)$ matches $\{W, \{w_1, w_2\}, \{w_1, w_3\}, \{w_1\}\}$, which is W-coherent. Therefore by proposition 129, $\langle W, R, T, F\rangle$ is not self-effacing.

$\langle W, R, T, F\rangle$ is not transitive, since Rw_2w_3, Rw_3w_2 but it's not the case that Rw_2w_2.

□

Proposition 133. *There are no values of T and F such that, all frame $\cap$-coherent and non-self-effacing accuracy frames with those values are euclidean.*

Proof. We appeal to the frame described in the previous proof. It is not euclidean, since Rw_2w_3, Rw_2w_3 but it's not the case that Rw_3w_3. □

Likewise:

Proposition 134. *There are no values of T and F such that all frame W-coherent and non-self-effacing accuracy frames with those values are transitive.*

Proof. Same as proof of proposition 132. □

Proposition 135. *There are no values of T and F such that all frame W-coherent and non-self-effacing accuracy frames with those values are euclidean.*

Proof. Same as proof of proposition 133. □

So even restricting our attention to those frames which are either frame $\cap$-coherent or frame W-coherent but not self-effacing, does not bring about introspection principles.

3 | Normative uncertainty and the accuracy argument for probabilism

Accuracy epistemologists represent the epistemic goodness of credences using functions known as *scoring rules*. Although there's widespread agreement on some of the constraints these scoring rules must satisfy, those constraints don't single out a unique scoring rule. There are well known concerns that the failure to identify a unique such function undoes the accuracy argument for probabilism, an argument that is central to many of the successes of the accuracy programme. This paper suggests that the failure of uniqueness in fact gives rise to normative uncertainty on the part of the ideally rational agents that accuracy epistemologists concern themselves with; uncertainty about the true theory of the epistemic good. Ethicists have discussed a number of 'uncertaintist' decision rules designed to guide agents uncertain about the true theory of the moral good. I show that epistemic analogues of the rules discussed in ethics allow accuracy epistemologists to avoid the problems that arise from uniqueness failures.

3.1 Introduction

This chapter asks what normative constraints hold when you are uncertain about the normative facts. It investigates this question in the context of a particular epistemological programme: the accuracy programme. As we have already seen, and somewhat roughly, the accuracy programme takes epistemic goodness to consist in setting your beliefs (in a suitably liberal sense of 'belief') close to the truth (in a suitably liberal sense of 'close to the truth'). Rational belief sets, on the accuracy picture, are those which are as close to the truth as possible (in a suitably liberal sense of 'possible').

This picture, suitably precisified, turns out to be quite powerful. As we have partly seen already, it has allowed proponents of the view to give novel arguments for a variety of foundational, structural epistemic norms. Those norms can be for outright belief (single-premise closure; pairwise consistency — and as I argued

previously, stronger requirements like multi-premise closure and consistency); or they can be for credences (probabilism; conditionalisation — though as I will argue in the following chapters, the story concerning this is not straightforward). In either case, and again, as to some extent we have already seen, the underlying arguments are the same. We define a *scoring rule* which measures how close a given belief is to the truth.[1] This provides a representation of how epistemically valuable a given belief is: in other words, an account of epistemic goodness. We then pick a *decision rule*, examples of which we saw in the previous chapters: for instance, 'you must maximise expected accuracy' or 'you must not be accuracy-dominated'. Lastly, we combine the two to show that if you want to maximise epistemic value, or avoid being accuracy-dominated, then you must be probabilistic, single-premise closed, and so on. As we have mentioned already, the programme has been highly successful, and used to examine not only foundational, structural epistemic issues, but more substantive questions about everything from epistemic akrasia to lying.[2]

But there's a problem. Certain foundational arguments (arguments for epistemic norms that adopt the value function-decision rule schema described above) rely on the selection of a unique scoring rule or accuracy function; on singling out a particular structure inherent to epistemic value. But all extant justifications of specific scoring rules fail to single out a unique structure in a philosophically satisfactory way. These justifications succeed in narrowing down the space of admissible accuracy functions, but they do not pick out one in particular. Absent a specific scoring rule, many accuracy epistemologists have worried that these ac-

[1]Previously we used 'accuracy function' to describe a function which measured the accuracy of an entire belief set. In this chapter it'll be convenient to have a way of measuring the accuracy of individual constituent beliefs or credences. The introduction of scoring rules is what lets us do that.

[2]For arguments for probabilism, see the introduction; for conditionalisation, see discussion in the following two chapters; for belief closure, see the previous two chapters; on akrasia, see Schoenfield (2016); on lying, see Krauss (2017).

curacy arguments no longer succeed. This is most clear in the case of the accuracy argument for probabilism, and it's this argument which will be our concern in this paper.[3]

I want to argue that things are not as hopeless as they seem. The reason is (I claim) that given the failure to axiomatise a unique scoring rule, the ideally rational agents that accuracy epistemologists concern themselves with are in a state of *normative uncertainty* — uncertainty about the true theory of the epistemic good. Once we recognise that this is the accuracy epistemologist's predicament, there are new moves we can make. For normative uncertainty is not unstudied — moral philosophers have for some time being trying to work out what you should do when you are unsure of the true moral theory. There are theories on the market. I am suggesting that versions of these theories are available in the epistemic case, too, and that they offer a way out for the accuracy epistemologist.

This way of doing things has another related consequence, besides blocking this objection. As we'll see, providing a philosophically satisfying axiomatisation that singles out one scoring rule in particular is very difficult. Once we invoke these uncertaintist rules, however, the burden is much lighter. All accuracy epistemologists need are two constraints that are widely agreed upon anyway.

In section 3.2 I lay out more clearly the problem that accuracy epistemologists face given the failure to successfully single out a unique accuracy function. In section 3.3 I explore a moral analogue of the same problem, and canvas some solutions which have been developed in the literature on *moral* uncertainty. In

[3]But similar questions might well arise for accuracy epistemologists who are interested in outright belief, as we saw in the preceding chapters. To recap: there are good reasons to think that the way to measure the accuracy of outright beliefs ensures that having true beliefs is good, and false beliefs is bad. And there are perhaps good reasons to think that the badness of false beliefs should outweigh the goodness of true beliefs (as we saw, for at least some kinds of accuracy coherence for outright belief this is the only way to ensure there aren't belief sets containing contradictory beliefs that are nonetheless undominated). But even these constraints dramatically underdetermine which accuracy function is the genuine one. I discuss this further in section 3.5.4 later.

section 3.4 I examine whether epistemic analogues of these solutions can work in the case of uncertainty about accuracy, and argue that they can. Section 3.5 draws out some additional consequences of our discussion. Section 3.6 concludes.

3.2 The problem

The problem we are interested in is known as the Bronfman problem.[4] To explain it, let me introduce Ira. Like all the agents we'll consider in this chapter, Ira has *credences* or *degrees of belief*. These degrees of belief are defined over a finite set of propositions $p \subseteq W$, and represented by a credence function Cr; a function from $p \subseteq W$ to the unit interval $[0, 1]$. Cr represents Ira's strength of opinion about the various propositions in $\mathcal{P}(W)$.

Ira is non-probabilistic, which is to say his credence function violates at least one of the axioms we gave in definition 12.[5]

Luckily for Ira, things are about to change. He's just picked up a copy of *The Accuracy Epistemologist's Almanac*. He now knows that he should care about having accurate credences — credences which are, loosely speaking, as close to the truth as possible given his information. Roughly, this means that the more credence you have in a truth, the better you are doing, epistemically speaking. But that by itself leaves open many different ways to measure the accuracy of credences. Reading further into the almanac, Ira becomes convinced that the correct way to measure accuracy is as follows:

First, the accuracy of a given credence is measured by a scoring rule, $s : [0, 1] \times \{0, 1\} \rightarrow [0, -\infty]$.[6] We denote this $s(Cr(p), p(w))$, where $p(w)$ is 1 if

[4]So named because it is first identified in an unpublished manuscript by Aaron Bronfman (Bronfman, n.d.); for very good discussion, see Pettigrew (2016a, ch. 5).

[5]In effect, though, we impose non-negativity on the credence functions by definition, as is standard in this literature.

[6]Typically scoring rules are defined as measuring *inaccuracy* rather than accuracy. For consis-

$w \in p$ and 0 otherwise. Since the scoring rules measure accuracy, higher scores are better, and a score of 0 is best.

The scoring rules recommended by the almanac have two important properties. First, they are strictly proper: that is to say, whenever your credences are a probability, your credences regard themselves as strictly better in expectation than any other credences.[7] Second, they are continuous.[8] We discussed strict propriety at length in chapter one; by way of a quick reminder, it is closely related to the thought that rational beliefs should be *immodest*; they should not regard some other beliefs as strictly better than themselves. If rational credences are to guide you, they cannot regard some other credences as a better guide, lest they undermine their own advice. Continuity is motivated by the idea that small shifts in strength of credence should correspond to small shifts in the (in)accuracy of those credences.

Scoring rules only allow us to measure the accuracy of an individual credence at a world. We measure the accuracy of a credence *function*, i.e. the set of all the agent's credences, at a world w, with an accuracy function A^c.

$$A^c(Cr, w) =_{\text{def}} \sum_{p \subseteq W} s(Cr(p), p(w))$$

When A^c is an accuracy function which measures the accuracy of Cr at w by summing $s(Cr(p), p(w))$ for all $p \subseteq W$, we say that A^c is *generated by* s.[9]

Ira is convinced by the almanac that the correct way to measure accuracy will

tency with other chapters, we take the negative value of those inaccuracy scores to give an accuracy score.

[7]We make this notion precise in the appendix. The reader might note that we talk of scoring rules, which evaluate the accuracy of credences, being strictly proper — rather than accuracy functions, which evaluate the accuracy of credence functions, as we did in chapter one. This is so we can stay close to the result from Predd et al. (2009) which we'll discuss shortly.

[8]More carefully, continuous in their first argument — again, we treat this properly in the appendix.

[9]We discuss this way of setting things up more in section 3.5.3.

satisfy the above constraints. Reading further, he gradually becomes close to convinced that the correct scoring rule, that is, the scoring rule which genuinely characterises epistemic value, is the *quadratic score*.[10]

Definition 136 (Quadratic Score)**.** $s_Q(Cr(p), w) =_{def} -(p(w) - Cr(p))^2$

With the almanac in hand, Ira realises that, in his current non-probabilistic state, he is guaranteed to be *accuracy dominated* given that the quadratic score correctly characterises epistemic goodness, for there is the following theorem about continuous and strictly proper scoring rules:

Theorem 137 (Predd et al.)**.** *Suppose A^c is generated by scoring rule that is strictly proper and continuous in its first argument. Then for all non-probabilistic Cr, there is some probability function Pr such that $A^c(Pr, w) > A^c(Cr, w)$ at all worlds. And for all Pr, there is no Pr' such that $A^c(Pr', w) \geq A^c(Pr, w)$ at all worlds, and no Pr' such that $A^c(Pr', w) > A(Pr, w)$ at some world.*[11]

What this theorem tells Ira is that, in his current non-probabilistic state, there is some probability function which is guaranteed to be more accurate than his own, *regardless* of which world he is at. Since Ira cares about being accurate, he should shift his credences to a probability function; having done so, he can no longer be accuracy dominated.

The line of reasoning Ira goes through has struck many epistemologists as attractive. But there's a hitch. The hitch is that not all scoring rules (accounts of epistemic value) agree on which probabilistic credence functions dominate Ira's credences. Moreover, not all rules *which satisfy the constraints we laid out above* agree on which probabilistic credence functions dominate Ira's credences. For

[10] As Gallow (2019, p. 5) explains, we need the quadratic score, rather than the commonly-invoked Brier score (Brier, 1950), since the former is defined for all propositions, while the latter is only defined for singletons $\{w\}, \{w'\}, \{w''\}$ etc. The difference doesn't matter if we assume probabilism; it does when we don't.

[11] Predd et al. (2009), Theorem 2, notation changed.

example, instead of the quadratic score, he might have considered the log score or the spherical score.[12]

Both of these scores are continuous and strictly proper. And therefore by Theorem 137, relative to each score, for any non-probabilistic credence function, there is some probabilistic credence function which dominates it and, moreover, is itself undominated. The problem is that both of these scoring rules can disagree with the quadratic score — and with each other — about which (probabilistic) credence functions are closer to credal perfection than Ira's.[13] Therefore they can disagree about which credence functions he seemingly ought to adopt.

And recall, although Ira was convinced by the almanac's case that the correct account of epistemic value was represented by a continuous and strictly proper scoring rule, he was only *pretty* convinced by the almanac's case for the quadratic score in particular. Ira is in a state of normative uncertainty, uncertainty about the true theory of the epistemic good.[14] Given this — given that he treats as live other candidate accounts of epistemic value (in particular, other continuous and strictly proper scoring rules) — it's not at all clear which scoring rule he should work with, and therefore it's not clear which probabilistic credence function he should move to, if any. For if Ira only cares about being as accurate as possible, and he doesn't know that adopting some probability function is guaranteed to make him any better off, it's unclear why he should do so. To make this more vivid, compare this familiar case[15]:

[12]See Pettigrew (2016a, ch. 5) for some discussion of these scores in connection with this issue.

[13]See Pettigrew (2016a, ch. 5) for a case of this kind.

[14]Can ideal agents like Ira be uncertain about true theory of the epistemic good? (Or moral good, see Hedden (2016, pp. 122-23) for discussion). I don't have a firm view on this. But note that, for our problem to even get going, it can't be that ideal agents automatically get this information. Since accuracy epistemologists *do* take this to be a problem, it makes sense to work on the assumption that, in fact, these agents can be normatively uncertain after all. Obviously, if they're not, the problem dissolves. This has some affinities with the normative externalist view we'll discuss later.

[15]Cases of this structure are well-known, see e.g. Parfit (1988); Jackson (1991). Pettigrew (2016a, pp. 71-73) similarly exploits the analogy in discussion of the Bronfman objection, more

> Ten miners are trapped during a cave-in, in either shaft A or shaft B. You know that all the miners are in the same shaft but you know not which. Moreover, the shaft is about to flood. You have to divert the water: either all of it into one shaft or split it between the two. If you put all the water into the empty shaft, no one will drown. If you put all the water into the occupied shaft, everyone will drown. If you split the water, half the miners will drown and half will survive.

The best outcome involves you correctly guessing the empty shaft. The outcome it is apparently rational to pick — splitting the water — is guaranteed to be worse than this outcome. After all, given this course of action, there will be deaths; if you guess correctly, everyone will live. But even though there is a better action available, it doesn't seem rational to pick it. It seems sensible to stick with an avowedly sub-optimal option. On the face of it, this is the situation Ira finds himself in. He knows that relative to the genuine account of accuracy, his credences are suboptimal. But since he doesn't know what the genuine account is, in cases where the scoring rules disagree it seems like he shouldn't shift from his non-probabilistic credences (Pettigrew, 2016a, p. 72).

Ira's predicament is analogous to the current state of play in accuracy epistemology. First, pretty much everyone is convinced that the correct scoring rule (the correct account of epistemic value) will be continuous and strictly proper.[16] That means that, relative to any particular scoring rule satisfying these constraints, any

on which later. Fitelson and Easwaran (2015) take the miners case as a motivating analogy when introducing their accuracy-theoretic approach to outright belief, which we discussed in previous chapters.

[16]The reader is referred to chapter one for discussion of strict propriety. Continuity is explicitly assumed somewhat less but (as I read things) that's typically just because it's unnecessary for certain accuracy arguments rather than because of any particular preference for discontinuity; the accuracy arguments for conditionalisation, for example, just exploits strict propriety. Given the restriction to finite W, we might also restrict the credence functions we're interested in to some finite subset of $[0, 1]$ in which case the continuity constraint is redundant.

non-probabilistic credence function is accuracy dominated by some probabilistic credence function, which is itself undominated.

What's more, some people are pretty convinced it's the quadratic score which correctly characterises epistemic value (the goodness of being close to credal perfection and the badness of being far away).[17] But extant attempts to single out the quadratic score, by imposing additional constraints upon the structure of epistemic value, have been philosophically unconvincing. Joyce (1998) faces complaints that its 'weak convexity' axiom cannot be justified (see Pettigrew, 2016a), Joyce (2009) faces further complaints over its 'coherent admissibility' constraint (see again Pettigrew (2016a) for criticism). The account given in Leitgeb and Pettigrew (2010a) is later rejected by Pettigrew (2016a), and the account given in Pettigrew (2016a) faces worries about its appeal to a 'decomposition' axiom relating accuracy and calibration: see Levinstein (2017a). So we can't rely on the quadratic score as the correct characterisation of epistemic value; other strictly proper scoring rules remain as live candidates to be the genuine representations of epistemic goodness.[18]

Because of this, it seems like Ira — and by extension, accuracy epistemologists — is (are) in an accuracy-theoretic Miners case: the current, non-probabilistic credences are guaranteed to be sub-optimal but there's no way to work out which option is better. This 'gap' is the Bronfman problem. The worry is, that on account of it, the accuracy argument for probabilism falls through.

One conclusion that some people have drawn from this failure to single out a

[17]Though see Levinstein (2012) for an argument for the log score. Pettigrew (2016a) defends the quadratic score explicitly, many other people appeal to the quadratic score in the literature though, as pointed out to me by an anonymous reviewer for a journal, it's unclear whether they do so out of conviction or convenience.

[18]A little more carefully: thus far, no one has convincingly singled out a unique scoring rule *up to positive affine transformation*. Singling out a unique function up to positive affine transformation would be enough, since those transformations preserve the intra-theoretical comparisons of the different possible credence functions by the scoring rule.

unique value structure is that it's *indeterminate* what the genuine scoring rule or accuracy function is. Pettigrew (2016a, ch. 5) treats the discussion of the Bronfman objection in this vein; Williams (2015) also discusses how indeterminacy interacts with the problem, though some of what he says might be interpreted as more sympathetic to the 'normative uncertainty' approach we're pursuing here. The 'Reply 1' in Richard Pettigrew's SEP discussion of the Bronfman objection also shares some affinities with the normative externalist picture we discuss below, though without the associated normative theory underlying the point (Pettigrew, 2019).

I don't have any decisive reasons for thinking about these matters in terms of normative uncertainty rather than indeterminacy, but I do think there are at least some considerations in favour of my approach. Firstly, since scoring rules are naturally interpreted as encoding a theory of the epistemic good, this way of setting up the problem is just quite a natural one. I would suggest that there is at least a tentative presumption in favour of thinking about things in these terms.

Second, and relatedly, in the moral case, the fact that an axiology doesn't single out a function that's unique up to positive affine transformation isn't obviously a reason to think *moral* value is indeterminate. It's not unheard of to draw that conclusion, but it's a little unusual (there is some appeal to indeterminacy in e.g. Broome (2004) and related critical range views). Instead, a more natural reaction to this failure is just that it's very difficult to give such an axiology. I think there's at least a reasonable case that something similar should apply in giving an accuracy-theoretic axiology.[19]

[19]For what it's worth, the accuracy-epistemological attempts to give a 'population ethics of belief' that I'm aware of also prescind from appeals to indeterminacy; see Carr (2015); Pettigrew (2018). Obviously, I don't think that the indeterminacy approach to population ethics is a non-starter, but I do think it would be philosophically a bit disappointing. Interestingly, it's a bit more common to invoke *incompleteness* in a population ethical axiology, e.g. where X and Y are two populations, and $\succeq$ is the 'better than' relation, but neither $X \succeq Y$ or $Y \succeq X$ — where this is a matter of determinacy, not indeterminacy; this is consistent with Carr (2015, p. 223).

Lastly, even if there are good arguments for modelling these Bronfman cases in terms of indeterminacy in the concept *accuracy*, it would still be interesting that (as I'll argue below) it's fruitful to approach this problem in terms of normative uncertainty. That in turn might give us abductive reason for thinking of the issue in normative terms rather than in terms of indeterminacy.

3.3 A moral analogue

I think the Bronfman problem can be solved. But to understand its solution, we need to briefly depart from epistemology. Let's introduce a second character: Una. Una is a highly conscientious and considerate person. However even highly conscientious and considerate people can be unsure of which theory of the moral good is correct, and this is the predicament Una finds herself in. Una is torn between two theories of the moral good in particular. She has high credence that desire satisfaction theorists have the correct account of wellbeing. But, despite this, she invests a small amount of credence in the truth of hedonism. One day, a scientist presents Una with the opportunity to get in an experience machine. It's a condition of entering the machine that she'd remain there for the rest of her life. If hedonism is true, then entering the machine is best. But if the desire satisfaction theory is true, then not entering the machine is best. In light of this uncertainty, what should Una do? (I hereby make a terminological stipulation that will operate for the rest of the paper: 'the good', and evaluative terms like 'best' will refer to the epistemic/moral good; deontic terms like 'should', 'ought' will be reserved for what agents are required to do *in light of* their normative uncertainty).[20]

[20]In the literature there's some debate about the different 'oughts' in play, my stipulation (I hope) lets us avoid these issues and any resulting terminological confusions (see Sepielli (2009); Harman (2015, pp. 70-71); MacAskill and Ord (2020, fn. 11) I'm not sure whether reserving the evaluative terms for the first-order theory presupposes a form of consequentialism. This is fine for our purposes since since accuracy views are often thought of as consequentialist, though see (Carr, 2017). See Greaves (2017, p. 12) for some discussion of the evaluative commitments of

Note how similar Una's predicament is to Ira's. She has multiple theories of the moral good that she treats as live. She wants to do as well as possible in the worlds she considers live.[21] But the theories of the moral good she leaves open disagree, and as a consequence, it's not straightforward how she should proceed.

Situations like the one Una finds herself in readily proliferate. Most people are not sure what the moral good consists in. For example, are the lives of humans and non-human animals equally valuable? If they're not, how great is the difference in their value? You don't have to be a moral sceptic to think such questions are difficult. It is not surprising, therefore, that moral philosophers have been interested in what is required of people in Una's situation. I want to look at three different ways in which we might respond to Una's predicament. These all take the form of views about our obligations under moral uncertainty. This class of views has a number of different names: they are sometimes known as metanormative rules, metadecision rules, or moral uncertaintist rules. I will use the latter. First, we'll see two kinds of internalist response. Then, a kind of externalist response. To clarify those labels somewhat, I stipulate that normative internalist treatments of normative uncertainty are internalist in the following sense: they make the morally required choice dependent on the agent's views about which moral theory is correct (in fact, the views we'll consider all make the choice depend on the agent's credences, but there's logical space for views that also invoke dependence on your outright beliefs, preferences, desires and so on). Normative externalist views are therefore those which do not make the morally required choice dependent on which moral theory the agent regards as correct.

It will also be useful to assign some credences to Una, and some value judgments of goodness and badness (i.e. some choiceworthiness judgments) relative

non-consequentialist theories.

[21]For ease of presentation, I'm excising any uncertainty about the non-moral consequences of her actions: she knows that if she enters the machine she will enjoy many pleasures but satisfy few desires, and if she doesn't enter, then she will satisfy many desires but enjoy far fewer pleasures.

to both the desire satisfaction view and the hedonic view.

	0.9 Credence	0.1 Credence
	Desire satisfaction	Hedonism
Enter	-200	100
Don't enter	10	10

Table 3.1: Una's Predicament

The numbers in the boxes represent how good and bad hedonism and the desire satisfaction theory regard entering and not entering the machine, respectively. By the lights of the desire satisfaction theory, entering the machine is very bad, whereas not entering is mildly okay. By hedonist lights, entering the machine is very good, whereas not entering is also only mildly okay.

I don't intend to spend much time discussing the strengths and weaknesses of the uncertaintist rules we'll consider. Our goal here isn't to resolve which theory of moral uncertainty is correct, but just to show that, regardless of which theory is correct, that theory can do some useful work for the accuracy epistemologist.[22] Instead, I'll just emphasise that the views we consider are the main ones in that literature (see MacAskill and Ord (2020) for discussion of all three). It's also worth noting that the three views do *not* exhaust logical space. That said, I think it's plausible that Maximise Expected Choiceworthiness and the Normative Externalist view exhaust the space of 'natural' views. I won't (and probably can't) argue for this with any sort of rigour but roughly: if you like 'internalist'-style normative views, you should like MEC; if you like 'externalist'-style views, you should like Normative Externalism, and the true theory must be either internalist or externalist (given the way in which I introduced them above). Hopefully the fact my argument works with any *natural* view, in the sense just described, is not

[22]In fact, when first thinking about this issue I'd wondered whether the ability to solve the Bronfman problem might serve as a useful test case for the viability of the various uncertaintist rules. But (as we'll see) the rules are all able to handle the problem, so it's not probative in that sense).

too poor a replacement for it working with any view whatsoever.

3.3.1 Normative Internalism

3.3.1.1 My Favourite Theory

The first kind of normative internalist response is known as the 'My Favourite Theory' view. Here is what it says.[23]

> **My Favourite Theory** You ought to treat as the good whatever any theory of the good in which you have greatest confidence says is the good.

So if you are more sure of hedonic theories of the good than any other theory, then you should treat hedonism as true. And if you are equally sure of hedonism and desire satisfaction theories, and more sure of those theories than in any others, you must treat either of them as true, though it is up to you to decide which. The idea is just that whichever theory (theories) you're most confident in should be the one to guide you in situations of normative uncertainty. For example, suppose you are 0.99 confident in Kantianism and 0.01 confident in virtue ethics. Faced with a moral problem, it would — on the face of it — seem bizarre to pay any attention to virtue ethics, given that you're almost certain it's a false moral theory.[24]

Intuitively, My Favourite Theory makes (what I take to be) the correct prediction in Una's case. Since she has more credence in the desire satisfaction theory than in hedonism, the theory predicts that she should not get in the experience machine, which — again on the face of it — looks like the correct prediction.

[23] See Gracely (1996); Gustafsson and Torpman (2014).

[24] In fact, given a 'belief is weak' view like those propounded in Hawthorne, Rothschild, and Spectre (2016); Dorst (2019); Rothschild (2020), My Favourite Theory just asks you to listen to whichever theory you believe is true — hardly an outrageous proposal.

3.3.1.2 Maximise Expected Choiceworthiness

My Favourite Theory has its detractors. An example worry is that it makes incorrect predictions in nearby variants of Una's case. Suppose Una had slightly different credences:

	0.49 Credence	0.51 Credence
	Desire Satisfaction	Hedonism
Enter	-200	100
Don't enter	10	10

Table 3.2: A variant Predicament

It is easy to see that My Favourite Theory now requires that Una enter the experience machine, for she is more confident in the truth of hedonism than the desire satisfaction theory, and given a hedonistic theory of the good, the right thing to do is get in the experience machine. But there is a clear sense in which Una is taking a big risk — after all, she is only slightly more confident in hedonism than the desire satisfaction theory, and if she turns out to be wrong, she will have done something morally very bad.

Problems like this one have prompted some normative internalists to look to other uncertaintist rules.[25] In particular, they have looked to a rule that blocks what goes wrong in this second case: that her decision procedure is insensitive to the risk of being wrong.

Let C be a function from moral theories $m \in M$ and options $o \in O$ to real-valued choiceworthiness scores. In Una's case, $M = \{D, H\}$ and $O = \{\text{enter}, \text{don't enter}\}$. From the decision tables above, we've seen (for instance) that $C(H, \text{enter}) = 100$ and $C(D, \text{don't enter}) = 10$.

We determine the expected choiceworthiness of an option by summing the choiceworthiness judgments of the various moral theories, weighted by how likely

[25]For some other problems, see MacAskill and Ord (2020, pp. 6-9).

the agent involved thinks it is that the theory is true.

Definition 138 (Expected Choiceworthiness)**.**

$$EC(o) = \sum_{m \in M} Pr(m)C(m, o)$$

We can now specify our second uncertaintist rule.[26]

> **Maximise Expected Choiceworthiness (MEC)**: You ought to pick an option o such that $EC(o) \geq EC(o')$ for all o'.

The motivation for a rule like this is something like the motivation for utilitarians who think you should maximise *expected* utility rather than actual utility. Just as empirical uncertainty can weigh on the rational thing to do, so, the thought goes, should moral uncertainty — and that uncertainty should be handled in a parallel way, namely in a way that is sensitive to the utilities you ascribe (or in this case, the choiceworthiness scores that are ascribed to) various options.[27]

This rule again generates the correct prediction in Una's first case: EC_{Una}(enter) $< EC_{Una}$(don't enter).[28] So Una is required to lie, which is intuitively the correct prediction. The rule also handles the second case, where it turns out that

[26] See MacAskill (2014); MacAskill (2016); MacAskill and Ord (2020) for defence.

[27] This also highlights something structurally unappealing about the My Favourite Theory rule. Consider the decision theoretic rule which told you to act as though you're in whichever state you think most likely to be live. That would be a terrible rule to live your prudential life by, as it would make you completely insensitive to possible losses and gains in states you consider (even slightly) less likely. A parallel problem is going to emerge for My Favourite Theory.

[28]

$$\begin{aligned} EC_{Una}(\text{enter}) &= Pr_{Una}(D) \cdot C(D, \text{enter}) + Pr_{Una}(H) \cdot C(H, \text{enter}) \\ &= 0.9 \cdot -200 + 0.1 \cdot 100 \\ &= -180 \end{aligned}$$

$$\begin{aligned} EC_{Una}(\text{don't enter}) &= Pr_{Una}(D) \cdot C(D, \text{don't enter}) + Pr_{Una}(H) \cdot C(H, \text{don't enter}) \\ &= 10 \end{aligned}$$

$EC_{Una}(\text{don't enter}) > EC_{Una}(\text{enter})$, again the correct prediction.[29] MEC is therefore predictively adequate, at least in the toy cases we've considered.

3.3.2 Normative Externalism

The two rules we've considered so far have been internalist; they've made the moral requirements depend, at least in part, on your credence in the various theories of the moral good. The last rule we consider is not internalist. It is externalist, for it makes the moral requirements insensitive to your credences, beliefs, etc. about whatever the true moral theory is.[30] In fact, the rule is very simple. It says:

> **Normative externalism**: You ought to treat as the moral good whatever is in fact the moral good.[31]

So if the desire satisfaction theory is the true theory of the good, that is the theory you ought to treat as true. If hedonism is the true theory of the good, that is the theory you ought to treat as true. This holds even if you are certain that hedonism is true when in fact the desire satisfaction theory is, or vice versa.

Remember, our goal here isn't to adjudicate between these various normative uncertaintist rules. So I don't intend to spend much time comparing the various

[29]

$$\begin{aligned} EC_{Una}(\text{enter}) &= Pr_{Una}(D) \cdot C(D, \text{enter}) + Pr_{Una}(D) \cdot C(D, \text{don't enter}) \\ &= 0.49 \cdot -200 + 0.51 \cdot 100 \\ &= -47 \end{aligned}$$

$$\begin{aligned} EC_{Una}(\text{don't enter}) &= Pr_{Una}(D) \cdot C(D, \text{don't enter}) + Pr_{Una}(H) \cdot C(H, \text{don't enter}) \\ &= 10 \end{aligned}$$

[30] Excluding certain pathological cases where, e.g. you're asked what moral theory you're most confident of, and the morally best thing is to tell the truth.

[31] See Weatherson (2019) for book-length defence; see also Harman (2015). Describing it as a decision rule might be a bit misleading: there's a sense in which what the normative externalist wants to say is that there are no special rules for reasoning under normative uncertainty whatsoever. We ignore these subtleties here.

rules, and I certainly won't be able to do justice to the various arguments that have been levelled in favour of, and against, normative externalism. But by way of motivation I note two things. First, it makes you genuinely responsive to the moral good, rather than what you *think* is the moral good — presumably an important feature of being a good moral agent.[32] Second, and relatedly, normative externalism holds fast in a case where MEC might seem to go awry. Imagine you are completely certain that anti-utilitarianism is true: you should always bring about the outcome with the worst expected utility. Then by the lights of maximise expected choiceworthiness, that is what you should do, for you give no weight at all to any other theories.[33] This is by no means an unanswerable objection to MEC, but presumably it's at least a little uncomfortable that there could be moral requirements of some stripe to maximise anti-utility. Normative externalism avoids problems like this, for it pays no attention to which theory you think is correct in issuing moral prescriptions; it instead enjoins you to be responsive to whatever in fact the good is (and since I take it anti-utility is not, in fact, the good, this problem is blocked).[34] I don't intend for these to be decisive reasons in favour of normative externalism, but hopefully they at least give a flavour of why normative externalism has been taken seriously.

3.3.3 Summary

Let's summarise. We've seen three views about how to act under conditions of moral uncertainty. All have received defence in the literature, although we've mostly side-stepped trickier questions about which view might be true. In the next section, we consider epistemic analogues of the rules we've just discussed.

[32] See Arpaly (2002) for classic defence of this view. See Sepielli (2016) for related discussion.

[33] Harman (2015) gives a version of this objection.

[34] See MacAskill and Ord (2020, fn. 11) for some discussion of what proponents of MEC might say in response to worries like this.

3.4 The solution

In this section, we see whether the uncertaintist rules we've just been discussing in the moral case can help us in the epistemic case. In both cases, the agents we are interested in are uncertain about the true theory of the good, whether that good be moral or epistemic. If the rules tell us how to proceed in the moral case, can they do so in the epistemic case, too? We'll now check the prescriptions of these uncertaintist rules and see whether they prescribe probabilistic credences to Ira. If they do, they give Ira a principled way of resolving his uncertainty in favour of probabilism.

3.4.1 My Favourite Theory

First, let's consider My Favourite Theory. We've assumed throughout that Ira is convinced (along with all other accuracy epistemologists) that the legitimate scoring rules are continuous and strictly proper. That's to say, he only has non-zero credence in those rules, while he has credence zero in the improper rules.[35] A bit more carefully: let p_s be the proposition that scoring rule s is the genuine scoring rule. We assume, then, that agents in Bronfman cases have, for some scoring rule, non-zero credence in the proposition that it is the genuine scoring rule, and that they *only* have non-zero credence in such propositions when the scoring rule is continuous and strictly proper. That's what it is to be in the case of interest: to treat some and only continuous and strictly proper rules as live or doxastically possible (which at least in the finite case, we can model by the lending of non-zero credence).

[35] Again, to defend this assumption a little more carefully: this is supposed to capture the state of play in accuracy epistemology, given the Bronfman objection. There's agreement that the genuine scoring rule is continuous and strictly proper, there's just less agreement about which rule in particular is genuine. The argument I give in this section won't go through once this assumption is relaxed; I discuss this further in fn. 39.

As long as this is so, then My Favourite Theory will enjoin Ira to work with the judgments of the scoring rule in which he is most confident. If there is no single theory in which he is most confident then it will permit him to pick. Either way, he will follow the judgments of a continuous and strictly proper scoring rule, and therefore, by theorem 137, he will end up with probabilistic credences.[36]

3.4.2 Maximise Expected Choiceworthiness

In the moral case, MEC combined the judgments of the various moral theories you treat as live, weighted by how likely you think they are to be the correct moral theory. We want a parallel notion in the accuracy case. First:

Definition 139 (Expected goodness at w of Cr' according to Cr)**.**

$$EG^{c}_{Cr}(Cr', w) = \sum_{p \subseteq W} \sum_{s: Cr(p_s) > 0} s(Cr'(p), p(w)) Cr(p_s)$$

(I use 'goodness' here instead of the mnemonic 'accuracy' to avoid confusion with more 'empirical' kind of expected accuracy we used when e.g. defining strict propriety earlier). $EG^{c}_{Cr}(Cr', w)$ returns the summed accuracy of Cr' at w according to the various scoring rules Cr treats as live, weighted by how likely Cr regards them to be the genuine scoring rule.[37] In this way, it is the accuracy-

[36]It's perhaps worth noting that My Favourite Theory comes close to what Pettigrew terms the 'subjectivist' response; agents can adopt any strictly proper scoring rule as they wish. Pettigrew claims that, given subjectivism, "Probabilism holds for agents whose attitudes towards inaccuracy are so rich and detailed that they specify a unique legitimate measure of inaccuracy" (Pettigrew, 2016a, p. 75), but worries there is no rational requirement for agents in Bronfman cases to have such a rich and detailed set of attitudes. In effect, My Favourite Theory fills this gap, giving agents a principled way of moving from the set of scoring rules open to them, to one rule in particular.

[37]Note that, somewhat unusually, the notions of expectation used in this section *don't* impose the requirement that Cr be a probability function, see Leitgeb and Pettigrew (2010a, pp. 214-15) for discussion of a similar notion, framed entirely in terms of the quadratic score. (Bronfman, n.d., p. 10) considers and rejects something like the strategy undertaken here (though the strategy is not framed in terms of normative uncertainty as such). In particular, he shares a worry of Joyce's that the concept of *expectation* doesn't even make sense for agents who aren't probabilistic. See

theoretic analogue of expected choiceworthiness at a world.

This notion only handles uncertainty about epistemic goodness at a world. We also want a notion which accounts for (empirical) uncertainty across worlds. We define the following notion of *expected expected goodness of a credence function Cr' according to Cr*:

Definition 140 (Expected expected goodness of Cr' according to Cr)**.**

$$EEG^c(Cr'|Cr) = \sum_{w \in W} Cr(\{w\}) EG^c_{Cr}(Cr', w)$$

I propose the above as analogous to the 'expected choiceworthiness' ethicists are concerned with.[38] It turns out that the following is true of expected expected goodness.[39]

Proposition 141 (Main result)**.** *If $Cr(\{w\}) > 0$ for some w, $Cr(p_s) > 0$ for some p_s, and for all s such that $Cr(p_s) > 0$ s is strictly proper and continuous in its first argument, then for all non-probabilistic Cr', there is some Pr such that $EEG^c(Pr|Cr) > EEG^c(Cr'|Cr)$.*

A full proof is given in the appendix, but by way of intuitive explanation: when you only leave open continuous and strictly proper scoring rules (and you do leave

Leitgeb and Pettigrew (2010a, esp. pp. 214-5) for discussion of this notion, which they in effect take to be the natural way of generalising the notion of expectations beyond the probabilistic case — I agree. See also Bronfman (n.d., p. 9) for discussion of this strategy.

[38] See Askell (2018) for related discussion. Note that we are now allowing empirical uncertainty where before it was suppressed — see fn. 21.

[39] See Askell (2018) for some thoughts on what happens when we relax the constraint that only continuous and strictly proper rules are live. Askell argues it is rationally permissible for agents to leave open some improper rules (Askell, 2018, pp. 305-306). We need not worry about these arguments, however, because the Bronfman objection involves cases where it is only continuous and strictly proper rules that are left open. Askell does note something like our result in passing (Askell, 2018, p. 301, fn. 18) (she cites the observation that any mixture of a strictly proper rule is strictly proper to Ben Levinstein) but not in enough detail to raise the issues we'll discuss below, with regard to leaving some proposition about a given scoring rule open, and some world open.

open some), it turns out that combining their judgments in the way expected accuracy does *also* yields a continuous and strictly proper scoring rule. And given Theorem 137, those rules will always regard probability functions as more accurate than non-probabilistic credence functions. So when you weight their combined judgments by how likely it is you're at various worlds, those weighted judgments come out preferring probability functions, too.

It is worth noting that the result assumes there is some w such that $Cr(\{w\}) > 0$, similar to Theorem 2 in Leitgeb and Pettigrew (2010b). There they just impose the requirement that there is some w such that $Cr(\{w\}) > 0$; though it may be possible to argue for it. Perhaps there is some tension in the notion of an agent who is at the start of their epistemic life, while also giving all worlds zero credence. For example, if W is supposed to be the set of worlds that are 'a priori' open, the agent should give them some non-zero credence, for what it is for a world to be open or live to an agent in this finite setting is for the agent to give it some non-zero credence. Relatedly, we might object that to treat no worlds as live is at odds with partitioning the space in the first place. Partitioning seems to require a grasp of the concept *possibility*, but then (in a finite setting) giving no world non-zero credence seems to suggest that one lacks that concept.

More promising, I think, is to appeal to the following:

Proposition 142. *If $Cr(p_s) > 0$ for some p_s, and for all s such that $Cr(p_s) > 0$, s is strictly proper and continuous in its first argument, then EG^c_{Cr} is generated by a continuous and strictly proper s.*

That is, when an agent leaves open (gives non-zero probability to) only continuous and strictly proper scoring rules, but possibly leaves open no worlds, then the combined judgments of the rules she leaves open will tell her that, given any non-probabilistic credence function, there is some probability function that is more accurate at every world. So a kind of dominance argument is still open even in

the case where no worlds are given non-zero probability. The argument would be that, if you are in a Bronfman case and have a non-probabilistic credence function, according to the combined judgments of the various scoring rules (theories of the epistemic good) you treat as live, you are accuracy dominated. This is not quite an epistemic analogue of MEC, but it is very close in spirit.

We've seen, then, that there are good reasons to think that the epistemic analogue of MEC can recover the accuracy argument for probabilism, so long as you are in a Bronfman case and only treat continuous and strictly proper scoring rules as live.

3.4.3 Normative Externalism

Let us turn finally to the normative externalist view. On this view, the scoring rule — theory of the epistemic good — which you should use is just whichever scoring rule is the genuine one. Since, by hypothesis, the genuine scoring rule will be continuous and strictly proper, theorem 137 ensures that in using this scoring rule, one will be probabilistic.

As mentioned earlier, Richard Pettigrew treats the Bronfman objection in terms of vagueness, rather than normative uncertainty. And our consideration of normative externalism as a response is close, in some respects, to his consideration of an epistemicist treatment of vagueness as a response (see Pettigrew, 2016a, pp. 71-72). In particular, Pettigrew notes that given an epistemicist theory of vagueness, and a vagueness-theoretic treatment of the Bronfman objection, we can construe Bronfman cases as cases where there is a determinate concept of accuracy, the agent cannot know what it is, but the agent is still in some sense beholden to whatever the true concept of accuracy is. This is similar to normative externalism, insofar as you may be ignorant about which theory of the good is genuine, but you are still in some sense beholden to it.

Pettigrew thinks we cannot charge Ira with irrationality for failing to shift to a probability function, on account of this vagueness. The reason Pettigrew gives is that, he claims, agents like Ira are effectively in a Miners Paradox case like the one we saw earlier. Recall, the lesson of the miners case is supposed to be that one can rationally do something that one knows is not best.

Given the similarities between the epistemicist response that Pettigrew rejects, and the normative externalist line being considered here, we should say why the two come apart. The short answer is that the normative externalist response falls out of a richer treatment of how we should react to the good. More specifically, there are good grounds (in my view) for thinking that what's required in variants of miners cases involving normative uncertainty is not what's required in variants involving empirical uncertainty. Here is an example from Brian Weatherson. It's hard to describe the case better than he does, so I quote at some length (Weatherson, 2019, pp. 33-34):

> Marilou is twelve weeks pregnant, and lives in a state where abortion is criminalized and, on occasion, heavily punished. Marilou deeply desires to have an abortion. Marilou is reasonably well-off, and as is the norm in states that criminalize abortion, reasonably well-off people are able to obtain abortions with a little assistance. Marilou asks her friend Shila for such assistance. Shila now has to make a choice. Shila is torn between two moral views about abortions twelve weeks into pregnancy. According to one, the potential that the fetus has to develop into a fully functioning human being means that aborting it is the moral equivalent of murder. According to another, the fetus has little or no moral standing on its own, so the importance of Marilou's autonomy means that Marilou should be able to get an abortion, and her friends should assist her in avoiding the oppressive laws against

abortion. Shila now has three choices.

> A. Assist Marilou in getting the abortion, which is either a way of respecting Marilou's autonomy and honoring their friendship, or is a way of being an accomplice to murder.
>
> B. Report Marilou's plans to the authorities, which is either horribly disrespectful to Marilou and a gross violation of their friendship, or bravely preventing a murder. (Assume that Shila knows that although the authorities aren't maximally vigilant about preventing abortions, they are obliged to act on incriminating information, so this tip-off will lead to Marilou's imprisonment.)
>
> C. Do nothing, suspecting that without her help, Marilou will carry the child to term and quietly adopt it out.

In either w_1, the world where abortion is permissible, or w_2, the world where it is not, C is bad. In w_1, Shila is a bad friend, and is tacitly collaborating in state oppression. In w_2, Shila is not taking simple steps that would remove the mortal danger facing an innocent human. But option C isn't catastrophic in either world. In w_1, Shila is not personally stopping Marilou getting an abortion, she just isn't helping Marilou break the law. (You can be a good enough friend and still draw the line between helping one move houses and helping one move bodies.) And in w_2, she's not killing anyone, or even letting someone be killed, just not being maximally vigilant in preventing a killing. So the case has the structure of a [miners] case.

And yet there is little to be said for C. The situation calls for moral bravery, one way or the other. (I think in the direction of A, but it

> doesn't matter for these purposes whether you agree with that.) And C is moral cowardice. Unlike in the cases involving factual uncertainty, it doesn't seem at all like the safe, prudent, commendable option.

I agree with Weatherson's judgment. Fix the story so that abortion is permissible, or impermissible, and you fix your judgments as to which action is permissible. But in neither case is that action C. What this indicates is that there is an important asymmetry between normative and empirical miners cases — as we saw above, in the latter kind of case it's the parallel of action C that's intuitively correct. The weaker lesson is that in objecting to epistemicism (or in our case, normative externalism), it's far from obvious that it is legitimate to infer anything about normative miners cases from our judgments in empirical miners cases; and so it's far from obvious (by normative externalist lights) that Ira should not shift to a probability function after all. The stronger lesson is that, in fact, what Weatherson's case teaches us is that we are required *not* to carry out a 'middling' action like C. The epistemic analogue would be that, given normative externalism, we are required *not* to stick with our non-probabilistic credences.

Another feature of the normative externalist view worth commenting on concerns guidance. A common objection to the view, discussed by both normative internalists and externalists, is that compared to normative internalist views, it provides inadequate guidance.[40] I said above that the purpose of this essay was not to adjudicate between the normative uncertaintist rules, and that remains true. I also think it is unclear how 'guiding' a normative epistemic theory must be to be tenable (Hughes, 2021). However, since (in my experience) accuracy epistemologists often make fairly strong assumptions about the guidance of the theory, it's at least interesting think about whether the normative externalist view is open to one

[40]See Lockhart (2000, pp. 8-9); Sepielli (2009, p. 8); MacAskill (2014, p. 7); Weatherson (2019, pp. 36-37).

who endorses the starting points of our discussion.[41]

Note that dialectically it might be okay if in fact normative externalism provided inadequate guidance — or at least, inadequate guidance by the lights of the accuracy epistemologist — since the normative internalist views we have discussed would be open to such theorists, and those views also circumvent the Bronfman problem. That said, I think there are at least moderately good grounds for thinking we can endorse the accuracy framework *and* the normative externalist outlook — or at least, that doing so is not really a worse combination of views than endorsing the accuracy framework and a normative internalist view. In short, the reason is that it's not clear that the internalist views provide a level of guidance that the normative externalist view cannot.

First, the maximise expected choiceworthiness view succeeds only at the cost of solving what is known as the problem of inter-theoretic value comparisons. Recall some of the examples we discussed earlier. We know that given a hedonic value theory, getting in the experience machine is better than not. We know that give a desire satisfaction value theory, getting in the experience machine is worse than not. We even have some comparative information about how much better or worse, by the lights of each theory — as given in Table 3.3.

In that table, we made a substantial simplifying assumption. We assumed that the numbers are all comparable on a single scale. For instance, we have effectively assumed that the value of not entering the experience machine given hedonism is equal to the value of not entering the experience machine given the desire satisfaction theory. Similarly, we have assumed that the value of entering the machine

[41]These kinds of internalist (in a sense looser and broader than the 'normative internalist' terminology used earlier) assumptions are not always made explicit, but they are occasionally. For instance, Leitgeb and Pettigrew write that "[T]he agent cannot be assumed to know at which world she is evaluating inaccuracies. So, by Ought-Can, we require a conception of the minimization of inaccuracy toward which the agent is epistemically capable of aiming" (Leitgeb & Pettigrew, 2010a, p. 213).

given hedonism is ten times greater than the value of not entering given the desire satisfaction theory. However there is nothing in either of the theories to underpin those comparisons. All the desire satisfaction theory can provide is a ranking of actions as good or bad insofar as they satisfy desires; similarly for the hedonic theory. Neither gives any information about how to translate a particular score into the currency used by the other ranking. The problem of inter-theoretic value comparisons is just that: how to exchange the respective value judgments of the differing theories into a common currency.

Here is another way of putting this: we could multiply every number given by the hedonic theory by a thousand. That would preserve all of the intra-theoretic comparisons made by the hedonic theory, and ultimately that is all the information the hedonic theory gives us. However when we went to take the weighted average of the scores in order to maximise expected choiceworthiness, we'd have thumbed the scale in favour of the hedonic theory's judgments. To be able to genuinely maximise expected choiceworthiness, we need to not thumb the scale in favour of a particular theory.

But working out how to do this in general is extremely hard. Even if we can make some intuitive comparisons — a painless, undiscovered, unnoticed murder is presumably worse by Kantian lights than it is by utilitarian ones — it's hard to see how we can do so in a more systematic fashion. Our ethical inuitions just do not seem rich enough. The problem is even worse in the case of scoring rules: it's very hard to have any sophisticated intutions about the judgments that the scoring rules make, making it very hard to know whether or not we've thumbed the scales in favour of a particular scoring rule or not. In consequence, it's presumably very difficult to actually apply the maximise expected choiceworthiness rule in practice. Moreover, it's not clear that it's much easier than just working out what

the true moral (epistemic) theory is.[42]

The problem of inter-theoretic value comparisons suggests that we are not guaranteed perfect access to (as it were) the utility functions needed to make the required calculation to maximise expected choiceworthiness. It is a familiar point that we probably do not have perfect access to the (subjective) probability function needed to make such comparisons. This is because we probably do not generally have have perfect knowledge of our own mental states. Philosophically, and as we mentioned in the previous chapter, probably the most influential reasons for this come in the form of Williamson's anti-luminosity arguments, which purport to show that there are no (non-trivial) mental states such that we are always in a position to know we are in them (Williamson (2000, ch. 4); though see, as we mentioned earlier, Schwitzgebel (2006) for arguments against the reliability of introspection of a more empirical bent).

It can be hard — or perhaps, pessimistically, impossible — to have knowledge of the credence and utility functions needed to execute the normative internalist rules. It can also be hard — or perhaps, pessimistically, impossible — to have knowledge of the true theory of the moral or epistemic good needed to execute the normative externalist rule. It's not at all clear that we are guaranteed to be in a better epistemic position with respect to any particular internalist uncertainst rule, and so it's not at all clear that the internalist rules can, in general, be more helpful in guiding us than the externalist rules. The combination of the accuracy framework with normative externalism, in that respect, looks about as good as its combination with normative internalism.

[42]See MacAskill, Bykvist, and Ord (2020, ch. 5) for discussion of the inter-theoretic value comparison problem. Weatherson, the normative externalist *par excellence* writes that "Personally, I suspect it is one of the hardest problems in all of ethics. None of the remotely viable solutions to it seem either obvious to the lay actor, or easy to implement" (Weatherson, 2019, p. 113) . Interestingly, he does seem to think it is at least in principle soluble, though I confess I find his argument for this unclear.

3.4.4 Summary

Let's summarise. Epistemic analogues of all three uncertaintist rules defended in the literature can handle the Bronfman objection. All three can give an account of how one should reason under circumstances of normative uncertainty, such that one reasons oneself to probabilism. By extending existing accounts of how to reason under normative uncertainty, we can recover the accuracy argument for probabilism in the case of interest.

But note that we've done more than just respond to Bronfman's objection. We've considerably weakened the philosophical requirements of a scoring rule axiomatisation. Accuracy epistemologists don't need to provide a philosophically satisfying axiomatisation that singles out a unique rule. Instead, they can just motivate continuity and strict propriety — a far easier task — and invoke one of the uncertaintist rules we've considered. This is enough to give the accuracy argument for probabilism.

3.5 Further issues

In this penultimate section I want to discuss some other issues raised by the foregoing.

3.5.1 Uncertainty about uncertaintist rules

The argument I've given has been that: assuming any of the decision rules for reasoning under moral uncertainty in the literature are true, then using that rule in a Bronfman case will recover the accuracy argument for probabilism. But we might wonder: what if the agent isn't sure which of these uncertaintist rules is the correct one? Does the argument I've given still go through in this case?

Let's work through these higher-order rules in turn. The first, a higher-order version of My Favourite Theory, would say: act according to any of the uncertaintist rules in which you're most confident. Since the rules we've seen will all lead the agents we're interested in to probabilism, this higher-order rule will do so, too.

Next, the higher-order analogue of MEC. Here, similar remarks apply to the first-order rule: every first order decision rule returns (one way or another) a continuous and strictly proper rule. So their sum (weighted by the higher-order rule) will return a continuous and strictly proper rule, too — leading one back to probabilism.

Lastly, normative externalism. This rule won't even countenance higher order uncertainty, it will just return the verdict it returned in the original case: namely, that the agents in Bronfman cases should respond to whatever is in fact the true theory of the epistemic good.

We can iterate all of these arguments up further levels of uncertainty.[43]

3.5.2 Other accuracy arguments

Another worry we might have about the foregoing is that it only considers one of the many accuracy arguments, not all. Maybe the theories of normative uncertainty we've considered can salvage the accuracy argument for probabilism in a Bronfman case. But if the cost is that none of the others go through in Bronfman cases, that is not much help. There are many accuracy arguments, and I can't speak to them all here. But I will note that there is at least one accuracy argument that is preserved in our case of interest. That is the accuracy argument for conditionalisation. This argument only relies upon strict propriety; it shows that if the epistemic good is represented by a strictly proper scoring rule, then the update procedure which maximises expected accuracy will be conditionalisation

[43] See MacAskill et al. (2020, pp. 30-31) for good discussion of related issues in the moral case.

(see Greaves & Wallace, 2005). Now suppose Ira is in a Bronfman case and wants to update. My Favourite Theory will tell him to pick any theory of the good in which he's most confident. Since these are all strictly proper, they will all tell him to conditionalise. The analogue of the maximising expected choiceworthiness rule will tell him to weight the judgments of the various rules he treats as live by how confident he is they are the genuine rule. As we've seen, their combined judgments are also a strictly proper rule, so they will also tell him to conditionalise. Lastly, the normative externalist 'rule' will require following the true theory of the epistemic good, which we've assumed is strictly proper — again, Ira is led to conditionalise. At the very least, then, we can recover the core Bayesian commitments of probabilism and conditionalisation.[44]

3.5.3 Non-additive rules

When setting out our framework earlier I made an assumption that was on the face of it is so innocuous it did not need naming. It was *additivity*, the condition that the accuracy of a credence function is just the sum of the accuracies of the constituent credences.[45] The reason sometimes given for endorsing additivity include that it is a very natural assumption, which makes the accuracy of a credence function nothing over and above the accuracies of the individual credences which comprise it (Pettigrew, 2016a, pp. 49-50).

More recently, there have been some investigations into non-additive accuracy functions, where the accuracy of a credence function need not be identical to the summed accuracies of its constituent credences (Pettigrew, 2022). Although I am not especially tempted to reject additivity, I do think it's interesting to look at

[44]In fact, we will question the accuracy arguments for conditionalisation at length in the next two chapters — but I bracket that for present purposes. Whatever update rule the competitor arguments end up endorsing, the key premise remains strict propriety, so we can give an analogous treatment of those, too.

[45]We called this 'separability' in chapter one.

whether permitting our agents to be unsure of (continuous, strictly proper) non-additive accuracy functions changes the argument I've given above in any interesting way. And as Pettigrew (2022, pp. 134–35) points out, there are apparently acceptable (strictly proper and continuous) ways of measuring accuracy that aren't additive (e.g. a rule described in P. J. Lewis & Fallis, 2019). Crucially, as Pettigrew shows, the accuracy argument for probabilism goes through *even with rules that are not additive* (but are still continuous and strictly proper).

It should hopefully be fairly obvious that, given this result, normative externalism and My Favourite Theory will again secure probabilism in the face of the Bronfman objection. The situation with MEC is slightly more tricky. First, we need to recall our earlier definition of a strictly proper accuracy function.

Definition 15 (Strict Propriety)**.** *A^c is strictly proper iff if Cr is a probability function, then for all Cr', $EA^{\mathbf{c}}_{Cr}(Cr) \geq EA^{\mathbf{c}}_{Cr}(Cr')$ with equality iff $Cr = Cr'$.*

Before we tacitly assumed this would hold for any accuracy function generated by a strictly proper s; now that we are no longer assuming additivity we will not want to assume this. Crucially, any weighted average of some strictly proper accuracy functions is itself strictly proper: in which case, the combined judgments of the strictly proper scoring rules, in the case of Maximise Expected Choiceworthiness, will always recommend a probability function.[46] So even if we relax the assumption of additivity, our argument goes through.

3.5.4 Outright belief

We can also consider how the lessons of this chapter bear on the framework of the preceding two. After all, just as we may be uncertain of the correct theory of the epistemic good for credences, we may be uncertain of the correct theory of

[46]The proof strategy of proposition 142 in the appendix is straightforwardly extended to strictly proper accuracy functions rather than strictly proper scoring rules.

the epistemic good for outright belief. This is particularly salient given that the constraints on the accuracy functions in this setting are so weak: besides requiring that $T > 0 > F$, the values of T and F seem more or less up for grabs.

It's fairly straightforward to see that, given My Favourite Theory or Normative Externalism, not much changes in these frameworks; the principles will just give one a way of singling out a particular accuracy function, and then everything will work as we saw already. MEC will be slightly different; one will take the weighted sum of the different candidate views for T, and for F. This means given $\cap B$-coherence, the uncertaintist rules will agree on the requirements of consistency and multi-premise closure, for they required those principles relative to any values for T and F. The requirements of the uncertaintist rules may come apart in the case of W-coherence, however. If you had most confidence in an accuracy function where $|F| < T$, or the genuine accuracy function was one where $|F| < T$, then W-coherence would require pairwise consistency. But it's consistent with both of those things that the weighted sum could turn out such that $T > |F|$, in which case maximising expected choiceworthiness would not require pairwise consistency after all.

3.6 Conclusion

We started out this paper with a problem: what happens when the attempts of accuracy epistemologists to single out the structure of epistemic value is only partly successful; what happens when agents are left with multiple continuous and strictly proper scoring rules, but those rules (can) disagree?

This paper has suggested a way out. We should appeal to uncertaintist rules used in ethics. Those rules give us a principled way of resolving uncertainty about what the correct account of the epistemic good is. Moreover, this strategy falls

out of a more general picture of how we should reason in conditions of normative uncertainty.

Much ink has been spilled trying to give a philosophically satisfying axiomatisation of a unique scoring rule. But given that any of these theories of normative uncertainty are correct, an axiomatisation that secures uniqueness is stronger than needed: it's enough to convincingly single out the continuous and strictly proper scoring rules. Many accuracy epistemologists will be satisfied this burden is already met.

Appendix 3.A Appendix

Definition 143 (Strictly proper scoring rules)**.** *s is strictly proper iff for all Pr, for all Cr, for all p,*

$$Pr(p)s(Cr(p),1)+(1-Pr(p))s(Cr(p),0) \leq Pr(p)s(Pr(p),1)+(1-Pr(p))s(Pr(p),0)$$

iff $Cr(p) = Pr(p)$, otherwise

$$Pr(p)s(Cr(p),1)+(1-Pr(p))s(Cr(p),0) < Pr(p)s(Pr(p),1)+(1-Pr(p))s(Pr(p),0)$$

.

Definition 144 (Mixture)**.** *s_M is a* mixture *of $s_1, ..., s_n$ iff $s_M = \lambda_1 s_1 + ... + \lambda_n s_n$, where $\lambda_1, ..., \lambda_n$ are non-negative constants and $\lambda_1 + ... + \lambda_n = 1$.*

The following result is well known:

Lemma 145 (Strict propriety and mixtures)**.** *For finite n, if $s_1, ..., s_n$ are strictly proper, any mixture s_M is strictly proper.*

Proof. Recall that s is strictly proper just in case

$Pr(p)s(Cr(p),1)+(1-Pr(p))s(Cr(p),0) \leq Pr(p)s(Pr(p),1)+(1-Pr(p))s(Pr(p),0)$

when $Pr = Cr$, with strict inequality otherwise.

Recall that $Pr(p)s_M(Cr(p),1)+(1-Pr(p))s_M(Cr(p),0) = \lambda_1[(Pr(p)s_1(Cr(p),1)+(1-Pr(p))s_1(Cr(p),0)]+...+\lambda_n[(Pr(p)s_n(Cr(p),1)+(1-Pr(p))s_n(Cr(p),0)]$.

Since s_1 through s_n are strictly proper, for $i \in \{1,...,n\}$,

$Pr(p)s_i(Cr(p),1)+(1-Pr(p))s_i(Cr(p),0) \leq Pr(p)s_i(Pr(p),1)+(1-Pr(p))s_i(Pr(p),0)$

when $Pr = Cr$, with strict inequality otherwise. So

$Pr(p)s_M(Cr(p),1)+(1-Pr(p))s_M(Cr(p),0) \leq Pr(p)s_M(Pr(p),1)+(1-Pr(p))s_M(Pr(p),0)$

when $Pr = Cr$, with strict inequality otherwise. So s_M is strictly proper. □

In fact we can strengthen this in a way that will be useful.

Definition 146 (General Mixture)**.** *s_M is a* general mixture *of $s_1,...,s_n$ iff $s_M = \lambda_1 s_1 + ... + \lambda_n s_n$, where $\lambda_1,...,\lambda_n$ are non-negative constants and there is some λ_i for $i \in \{1,...,n\}$ where $\lambda_i > 0$.*

Lemma 147 (Strict propriety and general mixtures)**.** *For finite n, if $s_1,...,s_n$ are strictly proper, any general mixture $s_M = \lambda_1 s_1 + ... + \lambda_n s_n$, where all λ_i, for $i \in \{1,...,n\}$ are non-negative and some λ_i is positive, is strictly proper.*

Proof. The proof of lemma 145 suffices to show this. □

The following property of continuous functions is well known.

Proposition 148. *For finite n, if $f_1, ..., f_n$ are continuous, then their sum $f_1 + ... + f_n$ is continuous.*

Lemma 149 (Main lemma)**.** *If $s_1, ..., s_n$ are strictly proper and continuous in their first argument, then any general mixture s_M is strictly proper and continuous in its first argument.*

Proof. Since $s_1, ..., s_n$ are strictly proper, by lemma 147 any general mixture s_M is strictly proper. Since $s_1, ..., s_n$ are continuous in their first argument, then by proposition 148 any general mixture s_M is continuous in its first argument. □

Proposition 142. *If $Cr(p_s) > 0$ for some p_s, and for all s such that $Cr(p_s) > 0$, s is strictly proper and continuous in its first argument, then EI_{Cr} is generated by a continuous and strictly proper s.*

Proof. Recall that:

$$EG^c_{Cr}(Cr', w) = \sum_{p \subseteq W} \sum_{s: Cr(p_s) > 0} s(Cr'(p), w) Cr(p_s)$$

Since all s such that $Cr(p_s)$ are strictly proper and continuous in their first argument, any general mixture $\lambda_1 s_1, ..., \lambda_n s_n$, where λ_i is replaced by $Cr(p_{s_i})$ is, by lemma 149, a scoring rule s_M that is strictly proper and continuous in its first argument. Then:

$$EG^c_{Cr}(Cr', w) = \sum_{p \subseteq W} s_M(Cr'(p), w)$$

So EA^c_{Cr} is generated by a scoring rule that is strictly proper and continuous in its first argument. □

Proposition 141. *If $Cr(\{w\}) > 0$ for some w, $Cr(p_s) > 0$ for some p_s, and for all s such that $Cr(p_s) > 0$ s is strictly proper and continuous in its first argument, then for all non-probabilistic Cr', there is some Pr such that $EEG^c(Pr|Cr) > EEG^c(Cr'|Cr)$.*

Proof. Since $Cr(p_s) > 0$ for some p_s, and for all s such that $Cr(p_s) > 0$, s is strictly proper and continuous in its first argument, then by proposition 142, EG^c_{Cr} is generated by a scoring rule that is strictly proper and continuous in its first argument. It follows by theorem 137 that for all non-probabilistic Cr', there is some Pr such that, at all w, $EG^c_{Cr}(Pr, w) > EG^c_{Cr}(Cr', w)$. But since there is some w such that $Cr(\{w\}) > 0$:

$$EEG^c(Pr|Cr) = \sum_{w \in W} Cr(w) EG^c_{Cr}(Pr, w) > \sum_{w \in W} Cr(w) EG^c_{Cr}(Cr', w) = EEG^c(Cr'|Cr)$$

. □

4 | Should you care about the way you look?

Most people would agree that it's good to be more informed rather than less. Presumably at least part of why we think this is because having more information helps you make better decisions. A famous result from Good (1967) shows (roughly) that acting after looking (viz. gathering more evidence) is better in expectation than acting without extra information. However, we don't normally think that gathering information is just practically useful. Having more information is epistemically useful, too. By itself, Good's result says nothing about the apparent epistemic value of looking. Instead, a closely related result due to Oddie (1997) shows (roughly) that evidence-gathering makes your opinions more expectedly accurate. If it's tempting to think that Good's result underwrites the practical benefits of evidence-gathering, it's equally tempting to think that Oddie's result underwrites the epistemic virtue of being more informed.

Oddie (like Good) assumes an internalist theory of evidence, according to which (roughly) your evidence is always perfectly informative about what evidence you have, or haven't, received. But epistemologists increasingly reject evidence internalism in favour of evidence externalism, according to which your evidence is not always perfectly informative about what evidence you have, or haven't received. Without the assumption of evidence internalism Oddie's result (and Good's, too) no longer hold — you can't always expect more information to be beneficial.

However, there's a wrinkle. Both Good and Oddie assume that you react to new information by conditioning on it. But according to accuracy epistemologists, it's at least questionable that you should react to new information by conditioning rather than using some other kind of update function, at least once evidence externalism is assumed. If other update functions maximise expected accuracy, we should test whether the evidence-gathering results hold on the assumption that we react to new information using them, instead.

We show that the evidence gathering theorems hold using these alternative, possibly expected accuracy maximising update functions. So whether evidence-gathering is worthwhile given accuracy epistemology and evidence externalism turns out to depend on exactly how the expected accuracy of update functions is measured, and which update functions you use as a result. In a slogan: you should care about how you look.

4.1 Introduction

Distinguish two questions about evidence: when should we seek it, and what should we do with it once we have it. The former question — one that goes back to at least Ayer (1957) — finds a famous reply in Good (1967), according to which (roughly) looking, i.e. gathering evidence, always maximises expected utility when doing so is cost-free (and so you are always permitted, and often obliged, to do so in such circumstances).[1] The latter finds a popular answer in the Bayesian norm of conditioning, according to which the agent's probabilities upon learning something should match the agent's prior probabilities, conditional on the evidence received.

Two (comparatively) recent epistemological innovations cast doubt on these replies. The first innovation is accuracy-first epistemology, which holds that epistemic policies are justified insofar as they promote accuracy. The second innovation is evidence externalism, understood here as the view that your evidence is not always perfectly informative about what your evidence is (or is not).[2] These innovations combine in a troubling way for the epistemology of evidence management. As Oddie (1997) shows, given evidence internalism — the view that your evidence is always perfectly informative about what your evidence is, or is not — gathering evidence (and then conditioning on it) always increases your accuracy in expectation. Similarly, as Good (1967) shows, given evidence internalism, gathering evidence (and conditioning) always increases the value of your actions in expectation. Yet once the assumption of evidence internalism is relaxed, look-

[1] Of course there are earlier instances of philosophers interested in the importance of evidence — Clifford (1877) being a notable, though still quite late — example. What's significant about Ayer is that he is (to the best of my understanding, anyway) the first to be interested in this question's connection with a background theory of (something like) evidential probability.

[2] There are other, closely related ways of understanding the internalist/externalist divide in epistemology, some of which will take the notion of perfect information about your evidence to be a corollary of the divide, rather than definitional of it. We discuss this in more detail later.

ing need be neither epistemically nor practically worthwhile. In fact, it will often be epistemically or practically forbidden. This might look like a disappointing result, for evidence externalists and for accuracy-first epistemologists. That gathering evidence is epistemically valuable seems almost platitudinous. Given the foregoing, it looks like it is not platitudinous, and is in fact false: bad news for the two epistemological innovations under discussion.

This paper argues, however, that the relationship between accuracy, externalism and evidence-gathering is more complicated still. On the one hand, while accuracy epistemology and evidence externalism combine to show that gaining new evidence and conditioning on it will often make you less accurate in expectation, not more, they also combine to show that the correct way to process new information may not be conditioning. Indeed, as Bronfman (2014) and Schoenfield (2017) show, arguments to the effect that conditionalisation maximises expected accuracy may have to assume evidence internalism. Once that assumption is relaxed, those arguments may not go through.

Which update function is most expectedly accurate turns out to be highly sensitive to how the expected accuracy of update functions is measured. Because of this, in testing whether gathering evidence always makes you more accurate (in expectation), we should assume that you will update using the rule that is best by the lights of accuracy and externalist epistemology. It turns out that once we grant evidence externalism, measuring the accuracy of update functions 'globally' (in a sense to be defined), then updating on new evidence using the globally most accurate rule is sufficient for versions of both Oddie's and Good's Theorems to go through. However measuring the accuracy of update functions 'locally' (in a sense to be defined), then updating on new evidence using the locally most accurate rule is insufficient for Oddie's, and Good's Theorems, to go through.[3]

[3]I will follow Ahmed and Salow (2019) in talking of 'counterexamples' to Oddie's and Good's theorems, or of those theorems not going through. To be clear, what's meant by this isn't a coun-

Whether evidence-gathering is vindicated in the accuracy epistemologist's framework depends very closely, then, on how you look. More generally, there is a question left open for accuracy-theoretic accounts of updating: can they secure the result that it's epistemically valuable to gather evidence?

Here is a roadmap for our discussion. In section 4.2 I introduce some background technical resources. In section 4.3, I discuss some important results that apparently show it is good to gather evidence. In section 4.4, we look at some circumstances under which those results don't hold. Section 4.5 points out that it's not clear why we should assume we'll *condition* on new evidence, and looks at some different accuracy-theoretic stories about how to update. Section 4.6 argues that, on at least some ways of measuring the accuracy of updates, if we respond to new evidence in the optimal way, it is still rational to gather new evidence. Section 4.7 concludes.

4.2 Background

Before we continue, I'll introduce some important background resources. Let W again be a (finite) set of possible worlds, and Pr a function from subsets of W to the interval $[0, 1]$ satisfying the constraints we described in the previous chapter. Informally, we can think of Pr as representing the agent's probabilities. For the sake of some concreteness, I'll assume in what follows that these are the agent's evidential probabilities. But I think everything I want to say would go through under the assumption these were the subjective probabilities of a fairly idealised agent.

Let E be a two-place accessibility relation defined on W. Informally, Eww'

terexample to the (proven) formal claim — the theorem is a theorem! — rather, it's a counterexample to the informal claim the theorem might be thought to encode, which is that gathering more evidence always makes your beliefs more accurate in expectation, or your actions better in expectation.

just in case for all your evidence says in w, you are in w'. We define $E(w)$ as the set of worlds w' at w such that Eww'. Informally, $E(w)$ is the strongest evidence you receive at w. We let $\mathcal{E} = \{E(w) : w \in W\}$. In terminology from a previous chapter, we say that $\langle W, E\rangle$ is a *frame*.

We will say that a theory of evidence is internalist iff E is an equivalence relation (and so E partitions W), and externalist otherwise. On an *internalist* theory of evidence, your evidence always tells you what your evidence is.[4] A theory of evidence is not internalist iff it is *externalist*, in which case your evidence does not always tell you what your evidence is.[5] E is an equivalence relation iff it is reflexive, transitive and euclidean iff $\mathcal{E}$ is a partition.[6]

Definition 150. *E is reflexive iff for all w, Eww.*

Definition 151. *E is transitive iff for all w, w', w'', if Eww' and $Ew'w''$, then Eww''.*

Definition 152. *E is euclidean iff for all w, w', w'', if Eww' and Eww'', then $Ew'w''$.*

We have given technical characterisations of 'internalist' and 'externalist'. But the technical characterisations correspond, at least with a good degree of approximation, to familiar, and more traditional, philosophical ones. We say that E is

[4]The label 'internalism' has many uses in epistemology, but I stipulate here that unless I say otherwise, by 'internalism' I just mean that your evidence is always perfectly informative about your evidence in the sense that, if p is part of your evidence, then it is part of your evidence that p is part of your evidence, and if p is not part of your evidence, it is part of your evidence that p is not part of your evidence.

[5]The label 'externalism' has many uses in epistemology, but I stipulate here that unless I say otherwise, by 'externalism' I just mean that your evidence is not always perfectly informative about your evidence in the sense described above: either there is some p that is part of your evidence, but it's not part of your evidence that it's part of your evidence, or, there is some p that is not part of your evidence, but it's not part of your evidence that it's not part of your evidence.

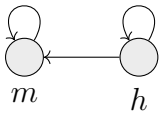
[6]More carefully, iff E is reflexive and euclidean; these suffice for E to be reflexive, symmetric and transitive, and therefore an equivalence relation. Since the euclidean and transitive properties will be important for our purposes, we use the slightly cumbersome characterisation given above.

internalist when it is an equivalence relation because internalists have typically (or at least, often) held that you have perfect access to your 'internal' mental states, such as the 'seemings', that constitute your evidence on many internalist views (see e.g. Pryor, 2000; Huemer, 2001; Smithies, 2019). Externalists deny that your evidence, whatever it is (for example, your knowledge, or beliefs formed via reliable mental processes, Williamson, 2000; Goldman, 2009) is the sort of thing you have perfect access to; something that can be reflected in E through the rejection of either transitivity or the euclidean property.

We will assume throughout that E is reflexive, and consequently, evidence is factive; if you receive the evidence that p, then p. This assumption is not completely uncontroversial, but it is popular, and has been defended in this literature.[7] Holding reflexivity fixed makes things a little cleaner, but otherwise we could lose the assumption if we wished. We will, however, question the other two properties required by internalist theories of evidence: the euclidean and transitive properties (note that although we question both of these properties, on our stipulated meanings, it is sufficient for a theory of evidence to be externalist that it rejects only one).

First, we repeat the externalist observation that there can be pairs of cases which, though (in some sense) internally indistinguishable may nonetheless differ in the evidence they provide. Take a case where the two possibilities m and h correspond to worlds where on the one hand, there is milk in the fridge,, and on the other, where there is an annoying milk hoax organised by your friend, where the milk in the bottle has been replaced with milk-coloured paint. Many think that such scenarios are captured by frames like the following, where an arrow from w to w' means that, for all your evidence says in w, you are in w':

[7] See e.g. Williamson (2000, pp. 201–202); Littlejohn (2014). See Pettigrew (in press-a) for interesting discussion of this matter in connection with accuracy arguments.



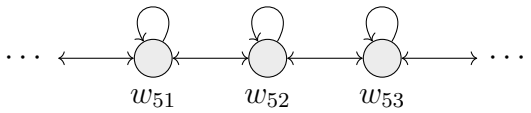
Even though (in some sense) the cases are internally indistinguishable, evidence externalists will think the evidence you receive in the two cases is different.[8] They will hold that if you see the milk in the fridge, you receive the evidence that there is milk in the fridge, and this evidence rules out the possibility that it is hoax milk; however, if you see the hoax milk, the evidence you receive does not rule out that possibility. One way (but certainly not the only way) to see why there might be an epistemically important difference between the two cases is to consider a reliability or safety condition on evidence (i.e. the sort of conditions that externalists about evidence are wont to endorse). In the case where there is milk, your belief that there is milk is safe and reliably informed; in the case where there isn't, that belief (plausibly) is neither safe nor reliably formed.

Cases like the above, described in terms amenable to evidence externalists, give rise to one kind of partitionality failure. Specifically, the frame is not partitional because E is not euclidean. This corresponds to the fact that in h, it is not evidence for you that there's milk in the fridge, but nor is it evidence for you that it's not evidence for you that there's milk in the fridge. Your evidence is uninformative, but it fails to tell you that it is uninformative. One gloss of this is that failures of negative introspection correspond to cases where your evidence is extremely misleading: so misleading that you have no hope of telling they are misleading. As Williamson (2000, p. 165) elegantly puts it, "[p]art of the badness of the bad case is that one cannot know just how bad one's case is".

These failures of *negative introspection* correspond to failures of E to be euclidean. Cases where E is not euclidean are perhaps the evidence externalist's

[8] See for instance Williamson (2000, ch. 8); Schellenberg (2013); Mitova (2015).

least controversial kind of partitionality failure.[9] But they are not the only kind. Here is another. Suppose you are looking at one of Williamson's (2014) "irritatingly austere" clocks, without markings for the different times. Since (plausibly) learning requires reliably getting it right, and the lack of markings makes for some unreliability, your evidence will not tell you exactly what time it is by looking. Rather, your evidence will tell you that the time is within a particular range:



At w_{51}, you receive the evidence that the time is between 50 and 52 minutes past, at w_{52} you receive the evidence that the time is between 51 and 53 minutes past, and so on.

These frames give rise to another kind of partitionality failure. Specifically, these failures of *positive introspection* arise because E is not transitive: you can receive some evidence without receiving the evidence that you've received it.[10] For example, for all the evidence you receive at w_{52} tells you, you are at w_{53} and have received the evidence that the time is between 52 and 54 past. Since you cannot rule out that you've received the evidence that the time is between 52 and 54 past, your evidence can't tell you that you've received the evidence that the time is between 51 and 53 minutes past. The claim that the evidential accessibility relation is not transitive is perhaps more controversial than the claim that the evidential accessibility relation is not euclidean. Nonetheless, many externalists are convinced that frames like the above provide the correct characterisation of inexact learning in cases like that of the unmarked clock.

[9] You sometimes find this sort of line being pushed by positive-introspection-endorsing evidence externalists, e.g. Das (in press, pp. 9-10, 19).

[10] See Williamson (2000, ch. 5) for the classic treatment.

Another piece of machinery we need is a fragment of decision theory. Let $\mathbb{R}$ be a set of outcomes; real-valued scores. The set O contains all functions from W to $\mathbb{R}$, representing options o. The outcome of an option o at a world w is $o(w) \in \mathbb{R}$.[11] We let $M \subseteq O$ be a *menu* of acts available to the agent (it needn't be the case that all options in O are things the agent can do). The expected utility EU of an option o relative to a probability function Pr is given by:

Definition 153. $EU_{Pr}(o) = \sum_{w \in W} Pr(\{w\})o(w)$

Lastly, let me fix some terminology regarding updating. By 'updating' I just mean any process through which you change your mind. The processes of interest in this chapter are encoded as functions, more on which below. I mostly talk of *receiving evidence* rather than *learning*. I do this because some of the update functions we will meet ask you, in some sense, to become certain of propositions that go beyond your evidence. I do not know whether to count such events as *learning* these propositions; to stave off this ambiguity, I will aim to talk mostly of what evidence the world has given you. Lastly, when I speak of the world giving you evidence, I mean that this is the strongest or total evidence it has given you. There is a sense in which, if the world gives you the evidence p, it has also given you the evidence (p or q). I here stipulate away any such ambiguity: when I talk about the evidence the world has given you, I refer to the strongest proposition it has given you (and implicitly assume there is only one such).

4.3 Good and Oddie

With this background in place, let's return to the philosophy of evidence-gathering. It's very natural to think that gathering evidence is a good thing to do. We nor-

[11] If you think a world fixes which option is taken, and so, for instance, $o(w)$ is ill-defined for w, o where o is not taken at w, replace worlds with *present states* or similar.

mally prize being more informed rather than less. People with more information seem to do a better job of navigating the world. They often have a clearer picture of things, and make better decisions.

These intuitive judgments can be underwritten with formal argument. Most famously in this vicinity, a result in decision theory due to Good (1967), and known variously as Good's Theorem, or the Total Evidence Theorem, says, roughly, that when gathering evidence is cost-free, you should always do so.[12] More carefully: given the standard decision-theoretic assumption that rational actions maximise expected utility, and given evidence internalism, gathering more evidence, conditioning on it and then acting will always have greater expected utility than acting without doing so — at least when the cost of evidence-gathering is negligible. This last qualification is important. If you have to pay a lot of money to acquire evidence, then it might be prudent to proceed without additional information rather than with. If you expect your evidence-gathering to unearth some traumatic truth, maybe it's prudent not to look — curiosity does kill some cats. Less dramatically, if you expect the evidence-gathering to be time-consuming, cognitively-exhausting, highly tedious, and so on, then it might not be worth the trouble. Good's Theorem is perfectly consistent with these observations. It says that if you assume that, in choosing whether to gather evidence or not, no particular cost is associated with the evidence-gathering, and evidence is internalist, then it will always pay to gather more evidence before acting.

Since it will be useful later, it's worth laying this out slightly more technically. The presentation here borrows heavily from Ahmed and Salow (2019). Let $\max_{o \in X}$ be a function selecting the $o \in X$ which maximises some quantity. The expected utility of acting given your current menu of options and current probabilities is therefore:

[12] Though inevitably, Ramsey got there before Good: see Ramsey (1990) for the idea in rough outline.

Definition 154.

$$EU(M, Pr) =_{df} \max_{o \in M} \sum_{w \in W} Pr(\{w\})o(w)$$

The expected utility of evidence-gathering corresponds to the following

Definition 155.

$$EU(look, Pr) = \sum_{E(w) \in \mathcal{E}} Pr(E(w)) \max_{o \in M} \sum_{w \in W} Pr(\{w\} \mid E(w))o(w)$$

Informally, $look$ is the option that corresponds to learning which member $E(w)$ of the partition $\mathcal{E}$ is true, updating on $E(w)$ and then choosing the option which maximises expected utility relative to this posterior probability function, where we assume that you update by:

> **Conditionalisation**: Upon receiving evidence p, your new probabilities Pr' should be your old probabilities Pr conditioned on p.

In general we will write Pr_p for the probability function conditioned on the proposition p. So conditionalisation says that upon receiving evidence p, your old probabilities Pr should become Pr_p.

Good shows that:

$$EU(look, Pr) \geq EU(M, Pr)$$

(with equality just in case there is some $o \in M$ that dominates all $o' \in M$ such that $o' \neq o$). A little briskly, the argument is as follows. The prior expected utility of acting is, given an instance of the law of total probability, and the fact $\mathcal{E}$ is a partition, equal to:

$$max_{o \in M} \sum_{E(w) \in \mathcal{E}} \sum_{w' \in W} Pr(\{w'\} \cap E(w)) o(w)$$

The expected utility of looking is equal to:

$$\sum_{E(w) \in \mathcal{E}} max_{o \in M} \sum_{w' \in W} Pr(\{w'\} \mid E(w)) Pr(E(w)) o(w)$$

Which is itself equal to:

$$\sum_{E(w) \in \mathcal{E}} max_{o \in M} \sum_{w' \in W} Pr(\{w'\} \cap E(w)) o(w)$$

Notice, however, that this very similar to just the expected utility of acting on your prior probabilities, but for the max operator taking scope over the first summand, rather than the other way round. And Good shows 'the sum of the maxima exceed the maximum of the sums' (Good, 1967, p. 320), and so in general,

$$EU(look, Pr) \geq EU(M, PR)$$

as required.

So, given the decision-theoretic assumptions just mentioned, and the assumption of evidence internalism discussed above, gathering evidence is always rationally permissible, and often rationally obligatory.

Good's Theorem gives an appealing first gloss explanation for the intuitive attractiveness of gathering evidence. It's good, because it makes us prudentially better off in expectation. This surely underwrites some of the appeal of gathering evidence. But if we reflect informally on the apparent goodness of evidence-gathering, it's very tempting to think that having more information is not just practically useful, but epistemically useful, too. While it's true that having more information typically helps you to make better decisions, it's also natural to think

it typically gives you a better picture of the world. By itself, Good's Theorem does not speak to that: it's only concerned with the connection between evidence-gathering and practical rationality. There's an obvious resemblance here to the accuracy argument for probabilism we discussed earlier. Dutch books and representation theorems gave reason to be probabilistic, but those reasons were, broadly speaking, prudential reasons. However, the requirement that your credences can be represented as a probability function seems like an epistemic requirement. The contribution of accuracy arguments was supposed to be that they identified an epistemic reason to be probabilistic, too.

We should ask, then, whether there are accuracy-theoretic reasons to want to gather evidence. On the face of it, the accuracy epistemologist should hope so. Gathering evidence seems like something that's epistemically good. On the accuracy-first view, epistemic goodness is grounded in accuracy. If accuracy epistemologists want to respect the judgment that evidence is epistemically valuable, they should have a story about the accuracy-conduciveness of evidence-gathering. Fortunately, there does seem to be such a story, as shown by Oddie (1997). Let us (re-)introduce an accuracy function A^c that measures the accuracy of a probability function Pr at a world w. Such functions are familiar from previous chapters. The expected accuracy of a probability function Pr' relative to a probability function Pr is given by

Definition 13 (Expected Credal Accuracy)**.**

$$EA^c_{Cr}(Cr') =_{def} \sum_{w \in W} Cr(\{w\}) A^c(Cr', w)$$

We recall the now familiar definition of strict propriety:

Definition 15 (Strict Propriety)**.** *A^c is strictly proper iff if Cr is a probability function, then for all Cr', $EA^c_{Cr}(Cr) \geq EA^c_{Cr}(Cr')$ with equality iff $Cr = Cr'$.*

As we have repeated *ad nauseam*, strict propriety has become a standard assumption in the literature on accuracy epistemology.[13] *Strictly proper* accuracy functions, as we have discussed repeatedly at this point, encode the requirement that your probability function should regard itself as epistemically best.[14]

As before, let $look$ be the option that corresponds to finding out which member $E(w)$ of a partition $\mathcal{E}$ of W is true, conditioning on $E(w) \in \mathcal{E}$, and choosing the probability function that maximises expected epistemic utility relative to this posterior probability function. That is,

Definition 156.

$$EA^c(look, Pr) = \sum_{E(w)\in\mathcal{E}} Pr(E(w)) \sum_{w'\in W} Pr(\{w'\} \mid E(w)) A^c(Pr_{E(w)}, w)$$

(We abuse notation, having let EA^c also take probability functions as arguments. It should always be clear from the notation which is meant). Oddie shows that — given evidence internalism — when evidence-gathering is epistemically cost free, the expected accuracy of looking is at least as great as the expected accuracy of not, with equality iff $\mathcal{E} = \{W\}$. Again, somewhat briskly: since $\mathcal{E}$ is a partition

[13] Oddie construes things in terms of epistemic *utility*, which is often thought of as a more general notion than accuracy. Typically, an epistemic utility function is regarded as an accuracy function just in case it's truth-directed, as we discussed in chapter one:

Definition 16 (Truth-directedness)**.** *If Cr and Cr' are credence functions, and $Cr(p) \geq Cr'(p)$ for all true propositions at w, $Cr(p) \leq Cr'(p)$ for all false propositions at w, and either $Cr(p) > Cr'(p)$ for some true proposition at w or $Cr(p) < Cr'(p)$ for some false proposition at w then $A^c(Cr, w) > A^c(Cr', w)$.*

The difference between epistemic utility functions and accuracy functions is irrelevant for our purposes because we assume both are strictly proper (see main text).

[14] See chapter 1 for discussion.

$$\begin{aligned}
EA^c_{Pr}(Pr') &= \sum_{w \in W} Pr(\{w\}) A^c(Pr', w) \\
&= \sum_{w \in W} \sum_{E(w') \in \mathcal{E}} Pr(\{w\} \mid E(w')) Pr(E(w')) A^c(Pr', w) \\
&= \sum_{E(w') \in \mathcal{E}} \left(\sum_{w \in W} Pr(\{w\} \mid E(w')) A^c(Pr', w) \right) Pr(E(w')) \\
&= \sum_{E(w') \in \mathcal{E}} EA^c_{Pr_{E(w')}}(Pr') Pr(E(w'))
\end{aligned}$$

Given this series of identities, the expected accuracy of not evidence gathering is:

$$\sum_{E(w) \in \mathcal{E}} Pr(E(w)) EA^c_{Pr_{E(w)}}(Pr)$$

Whereas the expected accuracy of evidence gathering is

$$\begin{aligned}
&= \sum_{E(w) \in \mathcal{E}} Pr(E(w) \sum_{w' \in W} Pr(\{w'\} \mid E(w)) A^c(Pr_{E(w)}, w') \\
&= \sum_{E(w) \in \mathcal{E}} Pr(E(w) \sum_{w' \in W} Pr(\{w'\} \cap E(w)) A^c(Pr_{E(w)}, w') \\
&= \sum_{E(w) \in \mathcal{E}} \left(\sum_{w' \in W} Pr(\{w'\} \mid E(w)) A^c(Pr_{E(w)}, w') \right) Pr(E(w)) \\
&= \sum_{E(w) \in \mathcal{E}} EA^c_{Pr_{E(w)}}(Pr_{E(w)}) Pr(E(w))
\end{aligned}$$

So evidence-gathering has greater expected accuracy than not iff

$$\sum_{E(w)\in\mathcal{E}} EA^{c}_{Pr_{E(w)}}(Pr_{E(w)})Pr(E(w)) > \sum_{E(w')\in\mathcal{E}} EA^{c}_{Pr_{E(w')}}(Pr')Pr(E(w'))$$

Given strict propriety, every term in the former summand, for evidence gathering, is strictly greater than the corresponding term in the latter summand, for not; finding out more information and conditioning on it promises more accuracy expectation than not — at least assuming that evidence internalism is true, and that the (epistemic) cost of gathering more information is negligible.

Even bracketing issues concerning evidence internalism, to be discussed shortly, it is not completely uncontentious that Oddie's Theorem can provide epistemic reason to gather evidence. For instance, we have — it is claimed — no reason "epistemic or otherwise" to count how many blades of grass there are in the meadow, even though that would confer on us lots of accurate beliefs (Greaves, 2013, p. 322).[15] As an attempt to block the possibility of such reasons, Greaves suggests restricting the sort of acts that accuracy epistemology (in her terminology, 'epistemic decision theory') can speak to, so that it only applies to purely epistemic acts like forming beliefs. In which case, as she points out, there cannot be accuracy-theoretic reasons to gather evidence, since gathering evidence is a non-epistemic act.

This is a reasonable worry. That said, I think there are probably ways around it, depending on how exactly we think about the evidence-gathering results. One is to claim that we do in fact have epistemic reason to gather evidence, and do in fact have epistemic reason to count the blades of grass — we just lack all-things-considered reasons to do these, or perhaps all things considered reasons to count the blades of grass. One thing this answer has in its favour is that Good's theorem

[15] We discussed some related issues in section 1.7.2.

apparently shows that we normally have prudential reason to gather evidence, too — but presumably it's not normally prudentially worthwhile for us to count blades of grass, because doing so is prudentially costly.

Another reply is to claim that, even if non-epistemic acts aren't things we can be moved to do by epistemic (accuracy-theoretic evaluations) alone, it's nonetheless an interesting result that — in a very general class of cases — those non-epistemic acts are evaluated positively anyway (something like this idea seems to be at work in, for instance, Dorst, 2020). For instance, maybe this tells us something important about evidence: that even if epistemic evaluations alone cannot, or should not, move us to collect it, it's nonetheless the sort of thing that, whenever collected, will make us better off. In sum, I think Oddie's theorem remains an important result, even if there are subtleties about how exactly it is to be interpreted.

I mentioned earlier that just as there were both prudential and epistemic arguments for the norm *probabilism*, there were both prudential and epistemic arguments for the claim that gathering more evidence is good. But, just as prudential arguments are of limited force when it comes to a norm that is distinctively epistemic — namely probabilism — I think they're also of limited force as concerns evidence-gathering. So, I claim, in the first instance we should be interested in whether evidence-gathering is accuracy-conducive, rather than just prudential-goods conducive. Or, at least we should if we endorse the theoretical framework of the accuracy epistemologist. Unfortunately for the accuracy epistemologist, it turns out that evidence-gathering is not always epistemically good for you after all, as we're about to see.

4.4 A failure of Oddie's Theorem

In this section, we provide an example of a case where it is not accuracy-conducive to gain more information.

4.4.1 Measuring accuracy

We've spoken somewhat abstractly about the accuracy of episodes of evidence-gathering. But to advance our discussion, it will be helpful to be more specific. For ease of presentation, I'll work with a particular, popular way of measuring accuracy: the quadratic score that we met earlier.[16]

Let $p(w)$ be an indicator function that maps w to 1 iff $w \in p$ and to 0 otherwise. Then we can give the quadratic score s_Q of a particular probability in a particular proposition as follows:

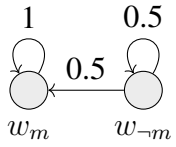
Definition 136.

$$s_Q(Pr(p), w) = -(p(w) - Pr(p))^2$$

The quadratic accuracy QA^c of a probability function at a world is defined as follows:

Definition 157. $QA^c(Pr, w) = \sum_{p \subseteq W} s_Q(Pr(p), w)$

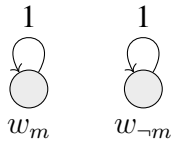
Lastly, we will enrich the frames we have been working with so far. Arrows from a world w to a world w' denote the probability you will give to $\{w'\}$ when you are at w. For instance:



[16] But everything we want to say relies on the fairly modest assumption that the accuracy measure is *strictly proper*, in the sense described above.

In this frame, the evidence you receive at w_m will lead you to be certain you are at w_m, whereas the evidence you receive at $w_{\neg m}$ leaves you completely uncertain as to whether you are at w_m or $w_{\neg m}$.

Having a way of measuring the accuracy of probability functions at a world lets us evaluate the accuracy of evidence gathering. Suppose you are antecedently unsure whether you are at w_m or not; you're 0.5 confident that you are and 0.5 confident that you aren't. Then given this frame, you can expect that looking will make you more accurate:



Since you are presently 0.5 sure you are at w_m, and 0.5 sure you are at $w_{\neg m}$, the expected quadratic accuracy (EQA) of your present probabilities is as follows.

$$EQA^c_{Pr}(Pr) = \sum_{w \in W} QA^c(Pr, w) Pr(\{w\})$$

$$\begin{aligned} QA^c(Pr, w_m) = QA^c(Pr, w_{\neg m}) &= -((1 - 0.5)^2 + (0 - 0.5)^2) \\ &= -(2(0.5)^2) \\ &= -0.5 \end{aligned}$$

But the expected quadratic accuracy of gathering evidence is the accuracy of your probability function upon learning you are at w_m, weighted by how likely it is you think you'll do so, plus the accuracy of your probability function upon learning you are at $w_{\neg m}$, weighted by how likely it is you think you'll do so. That is:

$$\begin{aligned} EQA^c_{Pr}(look) &= QA^c(Pr_{\{w_m\}}, w_m) \cdot Pr(\{w_m\}) + QA^c(Pr_{\{w_{\neg m}\}}, w_{\neg m}) \cdot Pr(\{w_{\neg m}\}) \\ &= 0 \end{aligned}$$

You expect in advance that looking will make you perfectly accurate. So you expect in advance that looking is better from the point of view of accuracy than not looking and sticking with your current probabilities — exactly as Oddie showed. Of course, that you expect evidence to make you more accurate does not guarantee that it will in fact do so, just as that you expect a decision to be prudentially beneficial does not guarantee it will be so. Declining to buy a lottery ticket sometimes does make you worse off, even if it didn't in expectation. What's important is that there's a systematic relationship between your perspective on epistemic value, and gathering evidence, that predicts gathering evidence is a valuable thing to do.

4.4.2 A counterexample

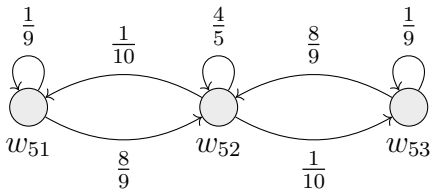
This systematic relationship, however, depends on the assumptions appealed to by both Good and Oddie. The first of those assumptions was that the cost of information gathering, whether that cost be epistemic or prudential, is negligible. In gathering evidence, you won't waste lots of time, spend lots of money, or fry lots of brain cells. We won't, for now, question this assumption. But as we saw above, there are good reasons to reject the assumption that evidence is partitional. Once we do, evidence-gathering need not maximise expected accuracy.[17] Instead, staying with your current opinions may do so.

That there are counterexamples to the evidence-gathering results is not a new observation; as mentioned, Ahmed and Salow discuss counterexamples to Good's

[17]Remember that we don't question reflexivity.

Theorem which exploit failures of partitional evidence (Ahmed & Salow, 2019), Das (in press) extends their investigation, while Dorst (2020); Das (2022); Salow (in press) discuss both failures of Good's Theorem and Oddie's Theorem in non-partitional cases. Nonetheless, it will be helpful to have a particular instance of the latter in mind.

Imagine you are in a case of transitivity failure. Specifically, we can advert to the example of Williamson's unmarked clock that we met earlier. To fix some details, let us suppose (a) you know ahead of time that the hand will be pointing at 51, 52 or 53 minutes past and (b), you are antecedently pretty sure it will be 52. Specifically, you are antecedently 0.8 confident the hand will point at 52, 0.1 confident it will point at 51 and 0.1 confident it will point at 53. Your posterior probabilities upon receiving the evidence $E(w)$ will be as follows:



Your expected quadratic accuracy, relative to those priors, is:

$$EQA^c_{Pr}(Pr) = \sum_{w \in W} QA^c(Pr, w)Pr(\{w\})$$

$$\begin{aligned} QA^c(Pr, w_{52}) &= -((1-0.8)^2 + (0-0.1)^2 + (0-0.1)^2 + (1-0.9)^2 + (1-0.9)^2 + (0-0.2)^2) \\ &= -(2(0.2)^2 + 4(0.1)^2) \\ &= -(0.08 + 0.04) \\ &= -0.12 \end{aligned}$$

$$
\begin{aligned}
&QA^c(Pr, w_{51}) = \\
&QA^c(Pr, w_{53}) = -((1-0.1)^2 + (0-0.8)^2 + (0-0.1)^2 + (1-0.9)^2 + (1-0.2)^2 + (0-0.9)^2) \\
&\qquad = -(2(0.9)^2 + 2(0.8)^2 + 2(0.1)^2) \\
&\qquad = -(1.62 + 1.28 + 0.02) \\
&\qquad = -2.92
\end{aligned}
$$

$$
\begin{aligned}
EQA^c_{Pr}(Pr) &= -(0.8 \cdot 0.12 + 0.2 \cdot 2.92) \\
&= -0.68
\end{aligned}
$$

The expected accuracy of looking and conditioning on what you learn, however, is:

$$EQA^c_{Pr}(look) = \sum_{w \in W} QA^c(Pr_{E(w)}, w) Pr(\{w\})$$

$$QA^c(Pr_{E(w_{52})}, w_{52}) = -0.12$$

$$
\begin{aligned}
QA^c(Pr_{E(w_{51})}, w_{51}) = QA^c(Pr_{E(w_{53})}, w_{53}) &= -((1-\frac{1}{9})^2 + (0-\frac{1}{9})^2 + (1-\frac{1}{9})^2 + (0-\frac{1}{9})^2) \\
&= -4(\frac{8}{9})^2 \\
&\approx -3.16
\end{aligned}
$$

$$
\begin{aligned}
EQA^c_{Pr}(look) &\approx -(0.8 \cdot 0.12 + 0.2 \cdot 3.16) \\
&\approx -0.728
\end{aligned}
$$

Since looking has a lower score than not-looking, you should expect that looking will make you more inaccurate. We have, then, a counterexample to Oddie's result that you can expect evidence-gathering to make you epistemically better off. Sometimes, when evidence isn't partitional, it need not.

4.5 Why condition?

On the face of it, such counterexamples are pretty surprising. As noted above, gathering evidence seems like the sort of thing that should be epistemically good for you. The counterexample appears to show that it's not — at least not in any systematic way that's likely to appeal to an accuracy epistemology. In fact, and for reasons I outlined partly above, I think there's at least a decent case that this counterexample should be more surpising and troubling than the counterexamples to Good's Theorem explored by Ahmed and Salow (2019). It's not very interesting that things which are epistemically good for us — as having more information would seem to be — can be prudentially bad for us. Memorising every entry in the dictionary would teach you a lot, but ruin your life. It is, I think, very interesting that something we thought was epistemically good for us can be epistemically bad for us after all. Relatedly, I think this kind of counterexample should be more troubling for accuracy epistemologists, who probably ought to be unconcerned (or at least, less concerned) about the practical value of evidence in the first place.

There are a few ways to react to the problem. One is to reject evidence externalism in favour of evidence internalism, since evidence internalism is sufficient for the recovery of Oddie's result.[18] We will not consider that here, for the reasons we discussed earlier.

Another is to claim that although the evidence-gathering platitude doesn't hold

[18]Though not necessary: see discussion of Dorst (2020) in main text below.

in generality, it holds *normally* or *usually* or *often enough* — basically, it holds universally in some restricted class of cases such that the evidence-gathering platitude strikes us as attractive. As a response, this isn't a non-starter, but it's disappointingly vague. How should we make 'normally' or 'often enough' precise? Is it guaranteed that once we do so, and focus on the normal or often cases, we'll be left with only those where the Good/Oddie result holds? It's far from clear that if you're in a normal case, you're in a partitional case.[19] While it might be pressed that failures of negative introspection are somewhat rare or unusual (though even that is questionable), inexact perception — dramatised with Williamson's unmarked clock — is something we presumably deal with on a day-to-day basis. Kevin Dorst has shown that the constraints we need to place on E for the Good/Oddie results to hold are, while short of partitionality proper, still fairly strong — they include transitivity, for instance, and therefore rule out (this way of modelling) inexact perception.[20] If we think you can be in a normal case while undergoing inexact perception, or that you undergo them even moderately frequently, then that will suffice to undermine this line of response.

There's a related response, which I think is a bit more promising. It attempts to salvage the evidence-gathering theorems by appeal to what we do when we ignore costly evidence gathering. The response I have in mind goes something like this. The evidence-gathering theorems are appealing despite being restricted. Sometimes evidence-gathering can be costly — epistemically or prudentially — and therefore sometimes it's irrational to gather evidence. However there's an important core of truth to the restricted claim made by the theorems, which is that it's rational to gather evidence in cases where doing so is not costly. This kernel of

[19] See Greco (2014); Carter (2019); Loets (2022) for applications of normality in epistemology, and in particular Loets for arguments that getting iteration principles out of a normality-based account will not be straightforward.

[20] Dorst (2020), see Cohen (2021) for a fascinating alternative to the 'orthodox' Williamsonian treatment of inexact perception.

truth is what vindicates our judgments about the goodness of evidence gathering. So far, this is just the standard treatment of costly evidence-gathering one finds in the literature. However, the response now argues that, analogously, that there's an important core of truth to the theorems with respect to partitional evidence; it's rational to gather evidence given that your evidence is partitional. Just as when we ignore costly evidence-gathering, this is what vindicates our judgments despite the possibility of non-partitional evidence-gathering.

I think there are at least a couple of ways to respond. One is to claim that we're interested in some idealised notion of 'pure' inquiry, on which it makes sense to abstract away from possible costs, but not from possible non-partitional evidence. A related gloss on this would be that costs are a fairly contingent feature of inquiring; non-partitional evidence is in some sense deeper, perhaps because it's a 'more' essential feature of creatures like us that we are not (in effect) omniscient.

Maybe a better way to go is just that the irrationality of evidence-gathering when doing so is costly is explicable, whereas the irrationality of evidence-gathering in cases of non-partitional evidence remains surprising. If you have to pay heavily to gain new information, it's unsurprising that it might not be worth doing so. But if you're in a non-partitional case and you should turn down free information, that *is* surprising. The reason for not 'idealising away' non-partitional evidence, then, is just that we can fairly easily make sense of the failures of the evidence-gathering theorems due to costs; we cannot readily make any sort of intuitive sense of the failures of the evidence-gathering theorems due to non-partitional evidence.

There are a couple more responses to the failures of the evidence-gathering theorems it's worth considering — and dismissing. We could claim that the counterexample involves a case where the act of evidence-gathering is costly, and therefore that it violates our stipulation that we only consider cases where evidence-gathering is free. But while evidence-gathering does turn out to be costly

in such a case (hence the counterexample) excluding it on such grounds seems to amount to ruling out counterexamples by fiat. Another reaction is to give up on the evidence-gathering platitude entirely. In that case, it would be good to have some explanation of why it seems appealing; an explanation that is more plausible than the one sketched above.

By far the best way to respond to the counterexample — I think, anyway — is to challenge an assumption it makes about how we update. We've assumed so far that when gathering information, you'll take new information you receive and condition on it. From the perspective of an agent concerned to maximise expected accuracy, it's not clear that this is what they should do. That's because it's not clear conditioning is the rational way to update, if you're concerned to maximise expected accuracy and you reject evidence internalism. In the literature on updating for accurate agents, it's unclear that given evidence externalism, conditioning maximises expected accuracy. Consequently it's unclear that conditioning is the way the agents we are concerned with would respond to evidence if they were evidence-gathering. Therefore it's unclear that the counterexample really is one, because it may assume agents would react to new information in a way that, if they were genuinely concerned to maximise expected accuracy, they would not.

To get an informal sense of why this might be problematic, consider the following sort of case from Steel (2018, p. 5):

> [S]uppose I have spent many years in an abusive relationship. I have finally found an opportunity to walk out, but have left some prized possessions behind at my former boyfriend's place. I want to get them back, but I'm afraid that if I return I'll see him. And I'm afraid that if I do he'll talk me into taking him back. I am extremely sure that leaving him is the right decision, but at the same time I know that he has, over the years, developed power over me, and that he understands

how to break down my resistance.

> Take the evidence here to be the content of the speech my ex-boyfriend is planning to give me. Here I have strong evidence that, when it comes to this particular piece of evidence, if I encounter it, then [...] even though nothing he says will present an actual reason for me to return, I will nonetheless be convinced by his speech to return. And this even if I go into the situation with a steely resolve, if I tell myself beforehand not to be fooled by him again, and so on, and even though the reasoning that I anticipate being faulty is of the fully explicit sort that we ordinarily classify as under our direct control

This is a case where the evidence-gathering slogan seems to fail. You can expect that gathering evidence in the case Steel describes will make you practically worse off, and given some pretty mild background assumptions, you can expect it to make you epistemically worse off too. But of the two stipulations we made, neither seem to obviously explain why evidence-gathering is bad in such a scenario. It's difficult to pin the blame on a failure of partitionality. We might reason that the content of the boyfriend's speech, for instance that 'you'll be happier if we get back together' is false, but we can set that aside by assuming that the evidence you receive is something true — there were many times that you were happy, for example (where the boyfriend neglects the times he treated you terribly).

Similarly, that evidence-gathering is bad in such a case is not obviously attributable to any costs incurred by the act of evidence-gathering, as in the examples we discussed earlier. There's no straightforward sense in which you pay to gather information in this scenario. Indeed, blaming this failure of the evidence-gathering platitude on costs looks close to again just ruling out counterexamples by stipulation.

The obvious diagnosis of what's going wrong in this case is that, upon receiving evidence, you're updating sub-optimally; you're reacting incorrectly to the information you've received — something that's hardly surprising given the unpleasant circumstances you find yourself in. For example, on a natural filling out of the case, even though you only get the information 'my boyfriend says I'll be happier with him', you become certain of 'I'll be happier with him', despite the latter not following from the former, and the latter not being supported (to that degree) by your evidence.

This informal example also provides the diagnosis we will explore in the remainder of this section. These purported counterexamples to Oddie's Theorem which exploit instances of partitionality failure may also be assuming that you will update sub-optimally on the evidence you receive. But if they do, it's not particularly surprising that such instances of 'evidence-gathering' can make you epistemically worse off, just as we've seen with Steel's case. What will emerge is that we need to test what happens in cases where we use the information we receive *optimally*. Is evidence-gathering worth our while in these cases?

Let's begin to fill out these vague remarks. There are (at least) two ways in the literature that epistemologists have evaluated the accuracy of different kinds of updates. We'll introduce these in turn.[21]

4.5.1 Evaluating updates globally

On the first way of assessing updates, we encode updates in *update functions*. These are functions f from a prior, Pr, and a world, w, to a posterior Pr', with the intended interpretation that, upon receiving evidence at w with prior Pr, you

[21] We won't, here, discuss some other ways philosophers have evaluated the accuracy of updates. We ignore the 'credal strategies' discussed by Briggs and Pettigrew (2020). The reason for this is that they assume evidence is partitional, putting them outside the scope of our inquiry. I comment on another approach we set aside in fn. 24.

will adopt posterior Pr'.[22] The evidence you receive at w is, as above, encoded in E; it is $E(w)$. For present purposes, we will also assume that if you receive the same evidence at two worlds, then the update function returns the same posterior probability at those two worlds.[23] That is, if $E(w) = E(w')$, then $f(Pr, w) = f(Pr, w')$. In the jargon:

Definition 158. $f \in \mathcal{F}$ *is* available *iff, if* $E(w) = E(w')$ *then* $f(Pr, w) = f(Pr, w')$.

Availability imposes a maximum level of responsiveness to evidence on the update function: if the same evidence is received in two worlds, the function cannot have differing responses in them. The omniscient update function $f(Pr, w) = Pr_{\{w\}}$ will thus be unavailable in many (semi-)realistic examples.

Since $f(Pr, w)$ is just a probability function, we can assess the accuracy of an update function at a world in the obvious way: $A^c(f(Pr, w), w)$. And with the accuracy of an update function at a world defined, we can assess the expected accuracy of an update function relative to Pr in the obvious way: by summing the accuracy scores at the different worlds, weighted by how likely it is you're at w according to Pr:

Definition 159.

$$EA^c_{Pr}(f) =_{def} \sum_{w \in W} A^c(f(Pr, w), w) Pr(\{w\})$$

(we abuse notation again, but context should resolve any ambiguities). As it were, we take the prior probability function's assessment of the likelihood of

[22] In fact, the formalism can be expanded beyond these 'deterministic' updates, to allow for the possibility that you are not sure how you will update given certain evidence (Topey, in press). We ignore such complications here.

[23] We relax this assumption later, and in particular, we do not require that this assumption is constitutive of evaluating updates globally.

being at various worlds. We will call this the *global* approach because it takes into account your opinions about the probability of every world when evaluating possible posterior probabilities.

Encoding things this way is very natural. Philosophically, the view is often undergirded by conceiving of the update functions as *updating plans* you can choose to adopt. The idea is that an agent concerned to maximise accuracy will adopt the updating plan which is most accurate in expectation. There are a number of interesting philosophical questions about this outlook: for instance, whether you are required — by some consideration of rationality or other — to actually carry out the plan you've adopted (see Pettigrew (2016a, ch. 14); Schoenfield (2016) for some discussion of the philosophical issues surrounding the planning framework). We will bracket these issues here and assume that, if an update plan maximises accuracy, then that's sufficient to justify that particular way of updating.

4.5.2 Local updates

The second way of assessing updates — the local approach — differs from the global approach in two key ways.[24] First, we no longer assess functions from a probability function and world pair to a probability function; instead, we just assess possible posterior probability functions directly. Let H be the proposition you receive as evidence. Second, then, on the local approach, the assessment is no longer straightforwardly by the lights of the prior probability function — which may give some probability to worlds not in H, the proposition received as evidence. Instead, we assess updates by considering which posterior probability

[24]There's another approach Leitgeb & Pettigrew consider, which is related to this one. Instead of only looking at the worlds in the evidence proposition H when calculating which posterior probability function maximises expected accuracy, we only look at those probability functions which give probability one to H and probability 0 to $\neg H$ — and see which of those maximise expected accuracy by the lights of the prior. We ignore this here since Gallow (2019, pp. 11-16) has convinced me that this way of justifying an update rule cannot yield conditioning, and so it is not relevant to our purposes.

functions maximise expected accuracy by the lights of the prior while only looking at those $w \in H$ when calculating the expectation.[25] That is, we select the posterior probability Pr' which maximises the following quantity.[26]:

$$\sum_{w \in H} Pr(\{w\}) A^c(Pr', w)$$

Because this way of evaluating possible posterior probabilities does not, in general, consider your opinions about the probability of all worlds, we will call this the *local* approach, in contrast to the global approach from before. How exactly this restriction to the worlds in H is philosophically motivated is a contested matter. Leitgeb & Pettigrew, who first introduced the technical features of this approach (Leitgeb & Pettigrew, 2010a, 2010b)) effectively take for granted that the effect of receiving evidence H is to rule out the $\neg H$ worlds. From an accuracy-first perspective, this is, as Gallow (2019) observes, a bit suspect. On the accuracy-first approach, rational belief transitions are supposed to be guided by accuracy, in which case, the ruling out of those $\neg H$ worlds ought to be so, too. But no accuracy-theoretic argument to that effect is forthcoming from Leitgeb and Pettigrew — it's effectively taken for granted that this is what receiving evidence does.

We might instead claim that this is a sort of 'arational' upshot of receiving evidence: something not apt for normative evaluation, and therefore not apt for evaluation as sufficiently or insufficiently accuracy-conducive. Gallow complains

[25] 'Expectation' here is being used extremely loosely, since the above is precisely not that so long as $Pr(H) < 1$. I'll continue to use this terminology, in absence of an obvious better alternative, but do bear in mind this is not strictly speaking an expected value.

[26] I'm slightly misrepresenting the Leitgeb and Pettigrew strategy here; they instead use a scoring rule — of the sort we saw in the previous chapter — to measure the (in)accuracy of the credence in each proposition. For ease of exposition I'm just assuming here that those inaccuracies can then be summed to yield the inaccuracy of the entire credence function. That assumption is not trivial, but it is widely accepted. As mentioned in the prior chapter, see Pettigrew (2022) for some interesting work on this.

that, if this is true, we're unable to judge apparently irrational responses to evidence as such — for instance, we are not able to judge as irrational becoming certain that the moon is made of cheese (ruling all out of the non-cheese-moon worlds) upon seeing that it's a rainy day.[27] I agree that it would be irrational to become certain the moon is made of cheese upon seeing some rain. I'm not convinced this shows that worlds being ruled out upon receiving a proposition as evidence must be a rational transition. As far as I can tell, Leitgeb and Pettigrew can explain our judgment of irrationality as follows. You receive the evidence that it's a rainy day. This rules out all of the non-rainy day worlds. You then, per their view, should pick the posterior probabilities which maximise expected accuracy. I want to defer discussion of what those probabilities are for the time being, but suffice it to say that for the Leitgeb and Pettigrew framework to be tenable, they must be reasonable posteriors (as their approach will in fact show them to be), and given even vaguely sensible priors, reasonable posterior probabilities will not then rule out all non-cheese-moon worlds. If this is right, then it's easy for Leitgeb and Pettigrew to explain the irrationality of ruling out non-cheese-moon worlds given the evidence you actually received: you've picked a posterior probability that doesn't maximise expected accuracy. All parties to the debate should accept that this is bad. In effect, then, I think Leitgeb and Pettigrew could locate the irrationality that Gallow is looking for in the posterior probabilities that are selected, rather than the worlds that the evidence rules out.

I do agree that it would be nice to have a firmer positive argument that the set of worlds one treats as live changes 'arationally' in response to new evidence, as seems to be needed on the Leitgeb and Pettigrew framework. It's not really clear why we should think this. But by the same token, I remain a bit sceptical that Gallow's negative point has much force. Gallow would presumably disagree, of

[27] Gallow (2019, p. 11). He credits the point to Jeffrey (1965).

course, and in fact Gallow (2019) defends the general Leitgeb & Pettigrew picture with a quite different background philosophical story. Instead of just assuming that the $\neg H$ worlds get ruled out when you receive evidence, Gallow treats receiving evidence as something that changes your epistemic values. Philosophically, the story is this. Upon receiving evidence H, your epistemic values change and you stop caring about accuracy at the worlds in $\neg H$: learning that H means you do not need to care about accuracy at $\neg H$ worlds. Technically, this is handled by having the new accuracy function — the function that reflects your new values — assign a constant value to all probability functions at $w \notin H$. That is,

$$A^c_E(Pr, w) = \begin{cases} A^c(Pr, w) \text{ if } w \in E \\ k \text{ otherwise} \end{cases}$$

where k is any constant. The expected accuracy of some possible posterior probability Pr', given the way your epistemic values change upon receiving as follows, is,

$$\sum_{w \in W} Pr(\{w\}) A^c_E(Pr', w)$$

The constant value has the following consequences: $\neg H$ worlds are irrelevant from the perspective of the expected accuracy calculation. Importantly, then, a posterior probability will maximise this value just in case it maximises the value Leitgeb and Pettigrew were interested in (Gallow, 2019, pp. 18-19).

Whether receiving evidence does enact this value change is up for grabs. There are, unsurprisingly, fairly stark disanalogies with the practical case. Some Dutch Book arguments effectively assume that it's always bad to act as though the $\neg H$ worlds are completely irrelevant. The Dutch Book argument for regularity, which we discussed in chapter one, takes something like this form. To repeat it, if you

give zero probability to some world w, then you would sell a $\$n$ bet that $\{w\}$ is true for $\$0$. This bet is guaranteed to make you no money at $w \in W \setminus \{w\}$, and lose you money at w, so the bet is weakly dominated by not selling any bet whatsoever — at least weakly dominated if we compare against all $w' \in W$ and not just $w' \in W \setminus \{w\}$. Insofar as you find this argument persuasive, you tacitly concede that your practical values are not like this; that you have ruled out some worlds (in this case, w) doesn't change how you value what happens at them. That's exactly how the above dominance reasoning is supposed to work.

Gallow doesn't say anything about whether learning should also change your prudential values. He probably should. It's quite natural to think that even when we've learned our lottery ticket has lost, a world in which we hold a losing ticket and get kicked in the head is not preferable to a world in which we have a losing ticket but no kick. That won't be the case if all acts at losing ticket worlds get constant values once we realise we didn't win the lottery. But his explanation of why learning changes your epistemic values isn't, in my view, rich enough to explain why it shouldn't change your prudential values too. This leaves him at least open to a sort of partners in guilt reply from the likes of Leitgeb and Pettigrew — or more generally, from someone disinclined to endorse this way of assessing the accuracy of updates.

A more serious problem, though, is that — as Gallow himself observes — on the accuracy framework, rational transitions are supposed to be grounded in accuracy. However this rational change in epistemic values, prompted by your learning that H, is not justified by means of an accuracy argument. It's also not clear how it could be.[28] Absent any story to that effect, it's not clear that accuracy epistemologists can adopt Gallow's view as it stands.

[28]Maybe this is a failure of imagination on my part, but insofar as it is, it's a failure shared by Gallow, who effectively concedes the point while maintaining that the Leitgeb & Pettigrew approach does no better (Gallow, 2019, pp. 17-18).

Fortunately, since the technical features of these two approaches are so similar — a posterior probability function maximises 'expected Leitgeb and Pettgrew accuracy' just in case it maximises 'expected Gallow accuracy', we don't need to settle this dispute here. For our purposes, what's important is that there's another way of evaluating the accuracy of updates.

4.5.3 Conditioning is accuracy-conducive (given certain constraints)

Assume evidence is partitional. Then whether you assess the expected accuracy of updates globally (Greaves and Wallace style) or locally (Leitgeb and Pettigrew, or Gallow, style), you will maximise expected accuracy (of the global or local variety) by conditioning. That is because the following results hold:

Theorem 160 (Greaves & Wallace, 2005)**.** *Assume $\mathcal{E}$ is a partition and if $E(w) = E(w')$ then $f(Pr, w) = f(Pr, w')$. Then the following quantity is maximised just in case $f(Pr, w) = Pr_{E(w)}$.*

$$\sum_{w \in W} Pr(\{w\}) A^c(f(Pr, w), w)$$

Theorem 161 (Leitgeb & Pettigrew, 2010b)**.** *The following quantity is maximised just in case $Pr' = Pr_H$.*

$$\sum_{w \in H} Pr(\{w\}) A^c(Pr', w)$$

Theorem 162 (Gallow, 2019)**.** *The following quantity is maximised just in case $Pr' = Pr_H$.*

$$\sum_{w \in W} Pr(\{w\}) A^c_H(Pr', w)$$

These results mean that whether we assess the accuracy of updating locally or globally, conditioning maximises expected accuracy when evidence is partitional. What about the case we are interested in, when evidence is not partitional? In reverse order: note that the local results makes no reference whatsoever to $\mathcal{E}$. They apply to any and all propositions $H \subseteq W$. Consequently, even in cases where evidence is not partitional, maximising local expected accuracy will lead you to condition.

The global result does, however, assume that evidence is partitional; it assumes that $\mathcal{E}$, the set of propositions $E(w), E(w')$ etc. partition W. By itself, then, the result says nothing about the case in which evidence is not partitional. Accordingly we cannot infer anything, from that result in itself, about the case where evidence is not partitional. In fact, it turns out that when evidence is not partitional, conditioning does not maximise expected global accuracy. But exactly what maximises expected global accuracy in its stead is a complicated question.

First, we note that we can encode the proposition that you have received p as evidence as the set of worlds w for which your total evidence there is p. That is, it is the set $\{w : E(w) = p\}$. We define $T(p)$ as the proposition that your total evidence is p, and $\mathcal{T}$ as the set $\{T(p) : \exists w : E(w) = p\}$; $\mathcal{T}$ is a partition.

Miriam Schoenfield then adapts the Greaves & Wallace framework to allow for the case of partitionality failure. On the intended interpretation, we think of the functions as update *rules* or *procedures*, the sort of thing a (doubtless highly idealised) agent could execute. Again, attention is restricted to those update functions that are *available* in the sense defined earlier. Schoenfield proves the following:

Theorem 163. *Assume that if* $E(w) = E(w')$ *then* $f(Pr, w) = f(Pr, w')$. *Then the following quantity is maximised just in case* $f(Pr, w) = Pr_{T(E(w))}$.

$$\sum_{w \in W} Pr(\{w\}) A^c(f(Pr, w), w)$$

This yields the update rule f sometimes known as 'meta-conditioning'.[29]

> **Meta-conditionalisation:** Upon receiving evidence p, your new probabilities Pr' should be your old probabilities Pr conditioned on the proposition that you have received p as evidence.

Updating by meta-conditioning means that, upon receiving evidence p, you condition on the proposition that you have received p as evidence. In the case where evidence is partitional, conditioning and meta-conditioning coincide. In the case where evidence is not partitional, the two come apart, and it is the latter which maximises expected accuracy given the constraints described above.

It's a contended matter that we should regard this as the expected-accuracy-maximising function on the global outlook, given partitionality failure. Gallow (2021) suggests two alternative ways of measuring the global expected accuracy of updates; these return correspondingly different maximising rules in the non-partitional case.

Both of Gallow's alternatives require thinking slightly differently about the update functions. In particular, they require thinking of them as *dispositions* to react to evidence. Since they are dispositions, they are not, like the *rules* or *procedures* above, something you execute. Rather, they are (or can be thought of as) automatic responses to receiving stimulus in the form of evidence (Gallow, 2021, pp. 491-92). Because of this, we no longer impose the availability constraint from earlier: we allow that there are two worlds, agreeing in the evidence provided but differing in the output of the update function. We do this because we allow that your evidence-responsive dispositions may '*misfire*' (Gallow, 2021, pp. 497-98); they may be likely to react one way to some given evidence, while leaving open a chance they will respond, on some occasions, differently. Compare (a familiar

[29] Hutchison (1999) and Schoenfield (2017) call it 'conditionalisation*'.

example) the disposition of a cup to break when dropped. Though cups are disposed to do so, most of us have had lucky occasions where a dropped cup survives intact. Similarly, though our update dispositions are cued to respond a particular way upon receiving certain evidence, we (are to) allow that they may misfire and react as though different evidence had been provided.

Thinking about things this way allows us to specify an alternative partition to the one above concerned with total evidence at a world, the partition: 'I've updated on p', 'I've updated on q', etc., for all $p \in \mathcal{E}$. We will notate these propositions as $U(p), U(q)$ etc, and write $\mathcal{U}$ for the set of them. Crucially, we allow that $\mathcal{T}$ and $\mathcal{U}$ may be distinct: in which case, we allow (at least) that there might be pairs of cases in which you receive the same total evidence but update in different ways. Sloganistically, you may update as though you had received total evidence you did not actually get.

Without the availability condition, there still needs to be some way of imposing restrictions on the set of update functions. We will want to rule out update functions like the omniscient update function described earlier, where whichever world you are at, you end up certain you are at that world. Without any such restrictions, these will be the update dispositions that maximise expected accuracy. But obviously, these are the sort of update dispositions that no agent of interest could possess.

The answer is, or at least may be, that the partition $\mathcal{U}$ is what settles which update dispositions are (in an informal sense) available. Since the cells of the partition correspond to the different ways the update dispositions can react to the evidence, it's extra-formally imposed on that partition that it could not correspond to the partition wherein every cell contains a single world. That partition cuts the space of possibilities too finely for any update dispositions to distinguish between.

Once this alternative partition is in place, Gallow uses it to argue, in two dif-

ferent ways, for competing update rules. These correspond to two different ways of measuring the accuracy of an update disposition.

According to the first way, we ask, for each cell of the partition, 'how accurate will my dispositions be if I update this way' (Gallow, 2021, pp. 499-500)? That is, we measure the accuracy of the update disposition as follows:

Definition 164.

$$EA'^{c}_{Pr}(f) = \sum_{U(p)\in\mathcal{U}} \sum_{w\in U(p)} A^{c}(f(Pr, w), w)$$

Gallow shows:

Theorem 165. *Assume that if* $w, w' \in U(p)$ *then* $f(Pr, w) = f(Pr, w')$. *Then the quantity* $EA'^{c}_{Pr}(f)$ *is maximised just in case* $f(Pr, w) = Pr_{U(E(w))}$.

The update dispositions which then maximise expected accuracy, at least with respect to the partition $\mathcal{U}$, are then as follows (Gallow, 2021, p. 500):

> **Update-conditionalisation:** Upon receiving evidence p, your new probabilities Pr' should be your old probabilities Pr conditioned on the proposition that you have updated as though you have received evidence p.

That is, according to update-conditionalisation, your new probabilities Pr' upon receiving evidence p should be $Pr_{U(p)}$.

According to the second way, we ask, for each cell of the partition, 'how accurate would my dispositions be if I updated this way?', taking into account now the possibility that the dispositions may misfire, and we may react as though we had received evidence unlike that which we actually received (Gallow, 2021, p. 499). Gallow suggests measuring the accuracy of updates this way:

Definition 166.

$$EA''^{c}_{Pr}(f) = \sum_{p \subseteq W} \sum_{w \in T(p)} Pr(\{w\}) \sum_{q \subseteq W} Pr(U(q) \mid T(p)) A^{c}(f(Pr, w), w))$$

Gallow proves:

Theorem 167. *Assume that if* $w, w' \in U(p)$ *then* $f(Pr, w) = f(Pr, w')$. *Then the quantity* $EA''^{c}_{Pr}(f)$ *is maximised just in case*

$$f(Pr, w)(\cdot) = \sum_{q \subseteq W} Pr(\cdot \mid T(q)) Pr(T(q) \mid U(E(w)))$$

.

In other words, you should update by what he calls:

Externalist-conditionalisation:

$$f(Pr, w)(\cdot) = \sum_{q \subseteq W} Pr(\cdot \mid T(q)) Pr(T(q) \mid U(E(w)))$$

With meta-conditioning, update-conditioning and externalist-conditioning on the table, we can now make the wrinkle in the counterexamples to Oddie's Theorem more explicit. Suppose you are concerned to maximise expected accuracy, and moreover, that you think the expected accuracy of updates should be assessed globally. The counterexample assumes you should calculate the expected accuracy of gathering evidence by weighting how likely it is you'll receive various pieces of evidence by how accurate you'll be having conditioned on that evidence — and summing the lot. But if you assess the expected accuracy of updates globally, then in the non-partitional case you should not plan to condition on the evidence you receive. To do so is to plan to use your evidence sub-optimally, from the perspective of accuracy-conduciveness. In which case — just as in Steel's

example above — the counterexample is not very surprising: it amounts to the result that gathering evidence and using it suboptimally can be worse than not gathering evidence at all. That is hardly a deep problem with the 'platitude' that evidence-gathering is epistemically valuable.

Our observations above let us make two key points. First, if we measure the expected accuracy of updates locally, then you will respond to new evidence by conditioning, and in cases of non-partitional evidence, it will sometimes be epistemically bad for you to gather evidence — as the counterexample we discussed in section 4.4 showed. Second, if we measure the expected accuracy of updates globally, then in cases of non-partitional evidence, you should not plan to update by conditioning, as that counterexample assumes. Therefore, it's an open question whether, given these expected accuracy-maximising functions, you should ever expect evidence-gathering to be epistemically bad for you. In the next section, we address this question.

4.6 Ensuring accurate updates

We've seen that, given non-partitional evidence, evidence-gathering can be expectedly epistemically bad for you, assuming you conditionalise. Whether you will conditionalise, given non-partitional evidence, itself turns out to be sensitive to how you evaluate the (expected) accuracy of update functions. If you assess them locally, you'll condition — and in the case of non-partitional evidence, the claim that it's good to gather evidence won't hold. The situation with global updates is less clear: we know you won't condition (you'll meta-condition, or use one of Gallow's rules, depending on exactly how the expected accuracy of updates is measured). But as a result, we don't know whether gathering evidence and updating on it in this way will preserve the value of gathering evidence. In

this section, we investigate this, by controlling for updates in the way suggested above: we suppose the agent will use the most globally expectedly accurate update function when gathering evidence, and investigate whether they can expect evidence-gathering to be, in general, epistemically valuable — or not.

Assuming we gather evidence and meta-condition on it, we can recover Oddie's theorem; it is always epistemically beneficial to gather evidence. The reason is that the expected accuracy of looking and meta-conditionalising on what you learn *just is* the expected accuracy of planning to meta-conditionalise. It is the expected accuracy of finding out which element $T(p) \in \mathcal{T}$ is true and conditioning on it.

The expected accuracy of not looking, however, is the expected accuracy of your present probability function; that is, the accuracy of your present probability function at each world, weighted by how likely it is you are at that world. That is, it is the expected accuracy of the trivial update function $f'(Pr, w) = Pr_W = Pr$:

But Schoenfield has shown that when f is the meta-conditionalising function, then

$$EA^c_{Pr}(f) \geq EA^c_{Pr}(f')$$

for all update rules $f \neq f'$ with equality just in case for all $w \in W$, $E(w) = W$.

That is to say, gathering more evidence and meta-conditioning on it will always make you at least as accurate in expectation as not doing so, and strictly more accurate in expectation when you think you may learn something non-trivial (formally, when there is some w such that $Pr(\{w\}) > 0$ and $E(w) \neq W$).

This is good news, and some evidence that the failures of Oddie's theorem we saw above were a consequence of handling evidence (from the accuracy-theorist's perspective) sub-optimally, rather than any more general feature of evidence-gathering.

Schoenfield's argument that meta-conditioning maximises expected accuracy relies upon the following formal result, 'extracted' from Greaves and Wallace

(2005).

Proposition 168. *For any partition of W, $P : \{P_1, ..., P_n\}$, consider the set of functions, $\mathcal{F}$, that assign members of P to probability functions. The member of $\mathcal{F}$, f, that maximises this quantity:*

$$\sum_{P_i \in \mathcal{P}} \sum_{w \in P_i} Pr(\{w\}) A^c(f(P_i), w)$$

is:

$$f(P_i, w) = Pr(\cdot | P_i)$$

when A^c is strictly proper.

The result itself does not say anything about the nature of the partition, and so — in principle — one may specify alternative partitions and use them in a similar way. This is what Gallow does when presenting 'update-conditioning', appealing to the partition $\mathcal{U}$ described earlier.

We can then offer a parallel argument to the one we gave in Schoenfield's case, replacing the partition $\mathcal{T}$ with $\mathcal{U}$. Gallow has shown that, given certain constraints, update-conditioning will be the most globally accurate update function in expectation.It follows again that it will be more accurate in expectation than the trivial update function that returns the same prior whatever evidence you receive. So, again, it vindicates Oddie's Theorem: looking is epistemically better than not. The reason is the same as it was in the case of meta-conditioning: we can redescribe looking in a given decision situation as the expected accuracy of update-conditioning on what you might learn at each world, weighted by how likely it is you are at that world.[30] And we can redescribe not looking in that

[30]See also Campbell-Moore and Salow (2020a, section 4.1) who make the same point, about (in their terminology) 'globally optimal' updates, in the context of the risk-sensitive agents.

decision situation as the expected accuracy of your current probabilities at each world, weighted by how likely it is you are at that world. Equivalently, we can redescribe not looking as the expected accuracy of conforming to the trivial update rule where at each world you condition on the tautology, weighted by how likely you think it is you are at that world. Gathering new evidence and update-conditioning on it is permissible in all cases, and required whenever you think you may learn something non-trivial.

We can run exactly the same trick for our final update function. If we gather evidence and then update on it using externalist-conditionalisation, then again we'll be more accurate in expectation if we expect to learn anything non-trivial; we can again recover Oddie's Theorem. The argument is the same as before: Gallow shows that externalist-conditionalision maximises expected accuracy; therefore it has greater expected accuracy than the update function which has you make no change in your beliefs, therefore (at least if you expect to learn something non-trivial) it's epistemically good for you to gather more information.

On the face of it, I think this is a promising result for the accuracy epistemologist. The diagnosis we suggested earlier for the failure of Oddie's theorem in these cases looks correct: the value of evidence theorem no longer held because you were using it, by your own lights, sub-optimally.

If these update functions recover Oddie's theorem, do they recover *Good's* theorem, too? This isn't without importance; for example, if the alternative globally most expectedly accurate update functions admit of failures of Good's Theorem, then there will still be tensions of sorts in the evidence-gathering platitude: it will be practically rational, but not epistemically rational, to ignore some evidence.[31]

So it's worth asking whether Good's Theorem holds if we uses these more expectedly accurate update rules. Recall that Good's theorem holds iff

[31]Though as Salow (in press, p. 4) points out, either one holding would be an interesting result.

$$\sum_{E(w)\in\mathcal{E}} Pr(E(w)) \max_{o\in M} \sum_{w\in W} Pr(\{w\} \mid E(w))o(w) \geq \max_{o\in M} \sum_{w\in W} Pr(\{w\})o(w)$$

Technically, nothing in the proof of Good's Theorem relies on the assumption that the partition is $\mathcal{E}$ — the set of propositions you might learn. All that matters is that $\mathcal{E}$ is a partition. This suffices to show that Good's Theorem holds if you update using meta-conditionalisation or update-conditionalisation, since these are just versions of conditionalisation restricted to a particular partition. Gathering evidence and updating on it by externalist-conditioning also has at least as great expected utility as acting without more information. The proof — which is effectively a version of Graves' (1989) proof that Good's Theorem holds for Jeffrey Conditioning — is a bit more long-winded so I relegate it to an appendix at the end of this chapter. These variant update rules not only recover Oddie's Theorem, then, but Good's Theorem too. If you update in an accuracy-conducive way, where the accuracy of the update is assessed globally, not locally, then not only will gathering evidence be epistemically good for you, it will be practically good for you, too.

4.7 Concluding remarks

Let's recap what we've found. Oddie's theorem says that, given partitional evidence, you can expect that gathering more evidence and conditioning on it will make you more accurate in expectation. Whether you should condition upon receiving non-partitional evidence is highly sensitive to how you assess the accuracy of updates: if you assess it 'globally', you won't, if you assess it locally, you will. We saw that given non-partitional evidence, receiving new evidence and condi-

tioning on it will not in general maximise expected accuracy. Consequently, if you evaluate updates locally, evidence-gathering will not always be a good thing for you to do: in some cases, you can expect that gathering more information and updating on it will make you worse off. If you assess the accuracy of updates globally, however, you won't condition upon receiving new evidence, because you'll think conditioning is not the most accuracy-conducive way to update. As long as you do update in the way recommended by expected accuracy, you will expect that gathering evidence and updating on it will make you more accurate — the evidence gathering platitude holds. What's more, if you update using what the global approach says is the most accuracy-conducive method, you'll also expect evidence-gathering to be practically useful — something the local approach cannot make sense of in cases of non-partitional evidence.

A natural way to interpret this result is that the global approach gets something importantly right about updating: it means that you'll react to evidence in such a way that evidence is valuable — something we might even want to treat as a data point.[32] That gathering evidence can be bad for you in even quite quotidian cases, according to the local approach, is on the face of it a mark against that way of assessing the epistemic goodness of updates. I think this is a tempting way to think about things. But unfortunately, I am not sure it is right. In the next chapter, we look at the update functions vindicated by the global approach in more detail, and suggest they get some important things about updating badly wrong, too.

Appendix 4.A Appendix

We show that, given you update by externalist conditionalisation, the expected value of gathering evidence and then acting is at least as great as the expected

[32] As Dorst (2020) does.

value of acting without new evidence.

The expected utility of acting without gathering new evidence is just the expected utility of whichever act maximises expected utility given your prior probabilities Pr.

$$max_{o \in M} \sum_{w \in W} Pr(\{w\})o(w) \tag{4.1}$$

To calculate the expected utility of evidence gathering, we need to know the expected utility of acting after you have updated with externalist conditionalisation. Suppose you have updated as though your evidence is p; or $U(p)$. Then the expected accuracy of your action, given this update, is:

$$max_{o \in M} \sum_{w \in W} \sum_{T(p) \in \mathcal{T}} Pr(\{w\} \mid T(p))Pr(T(p \mid U(p))o(w) \tag{4.2}$$

Since there are multiple possible ways that you will update open to you, the expected utility of gathering evidence, and then acting, is:

$$\sum_{U(p) \in \mathcal{U}} Pr(U(p))max_{o \in M} \sum_{w \in W} \sum_{T(p) \in \mathcal{T}} Pr(\{w\} \mid T(p))Pr(T(p) \mid U(p))o(w) \tag{4.3}$$

To show that the expected utility of gathering evidence and then acting is at least as great as the expected utility of acting without gathering new evidence, we show: (4.3) $\geq$ (4.1).

First, we observe that given the law of total probability, (4.1) is identical to

$$max_{o\in M} \sum_{w\in W} \sum_{T(p)\in\mathcal{T}} Pr(\{w\} \mid T(p))Pr(T(p))o(w) \tag{4.4}$$

Another application of the law of total probability yields

$$\max_{o\in M} \sum_{w\in W} \sum_{T(p)\in\mathcal{T}} \sum_{U(p)\in\mathcal{U}} Pr(\{w\} \mid T(p))Pr(T(p) \mid U(p))Pr(U(p))o(w) \tag{4.5}$$

Which is equal to

$$\max_{o\in M} \sum_{U(p)\in\mathcal{U}} Pr(U(p)) \sum_{w\in W} \sum_{T(p)\in\mathcal{T}} Pr(w \mid T(p))Pr(T(p) \mid U(p))o(w) \tag{4.6}$$

(4.3) and (4.6) are alike, but for the max operator taking scope over the summation of $U(p) \in U$ in (4.3), and the latter taking scope over the former in (4.6). The lemma from Good (1967, p. 320), which we discussed in section 4.3, shows that in this case: (4.3) $\geq$ (4.6), as required.

5 | Some hard truths about accurate updating

A core part of the hegemonic programme of *accuracy epistemology* has been to show that various epistemic virtues, though at first glance not grounded in the goodness of accuracy, can in fact be shown to be so. The more epistemic virtues that can be grounded in accuracy, the stronger the case for an *accuracy-first* epistemology according to which accuracy is the fundamental epistemic good. In this chapter we explore how this hegemonizing programme fares in the case of Bayesian updates. It's widely known that, given evidence externalism — roughly, the view that your evidence sometimes does not tell you what evidence you have — then the orthodox conditionalisation update function is not vindicated by considerations of accuracy. Instead, and depending on how exactly you measure the accuracy of the update functions in question, various conditionalisation-*like* rules are vindicated. You might think that that these variant rules are *close enough* to orthodoxy to count as a victory for the hegemonizing programme (and one sometimes gets the sense from hegemonizers that this is thought to be the case). In this chapter, we argue that this is not so. Although these update functions are in some respects similar to conditionalisation, they all violate certain important evidential strictures. Consequently, they all fail to properly hegemonize evidential features of epistemic rationality. Various ways out of the problem are assessed, with the most promising looking to be a partners-in-guilt style response to the effect that the trouble does not depend on accuracy epistemology in particular, but is instead part of a deeper tension between the idea that your beliefs should be self-recommending, and evidence externalism.

5.1 Introduction

According to some of its chief proponents, the programme of accuracy epistemology is a *hegemonic* one.[1] We take a theory of the epistemic good — accuracy — and we attempt to show that all other epistemic virtues can be explained in terms of it. As we've seen throughout this thesis, the hegemonic ambitions of the accuracy programme have been concerned firstly with structural coherence requirements like probabilism, but also evidential requirements. One important

[1] Chief among those chief proponents is Pettigrew (2016a).

task has therefore been to show that epistemically good ways of responding to evidence can be explained in accuracy-theoretic terms.[2]

The orthodox Bayesian theory of updating is conditionalisation. This theory encodes a number of independently attractive properties of responding to new evidence. As we have seen, there are a number of accuracy arguments for conditionalisation (Greaves & Wallace, 2005; Leitgeb & Pettigrew, 2010a, 2010b; Briggs & Pettigrew, 2020). However they often assume a theory of evidence — in particular, an *internalist* theory of evidence, in a sense we have discussed — that on philosophical grounds is very questionable. Once we use a philosophically adequate, externalist theory of evidence, then — as Bronfman (2014) and Schoenfield (2017) observe, and as we saw in the previous chapter, some of those accuracy arguments — specifically, those which evaluate the expected accuracy of updates globally — no longer go through. Instead, and depending on how exactly the expected accuracy of updating is measured, alternative conditionalisation-like ways of updating turn out to maximise expected accuracy (Schoenfield, 2017; Gallow, 2021).

It is an important question whether accuracy theorists really can account for various properties that properly responding to evidence seem to require. If they cannot, that would be an important limitation of the accuracy programme. Relatedly, one sometimes gets the impression that these variant update rules are thought 'close enough' to conditioning for the purposes of the accuracy programme's hegemonic ambitions. Whether they are in fact close enough remains to be seen.

In fact, I think these variant ways of updating cannot make sense of important parts of evidence-responsiveness. Therefore, to adopt them is to sacrifice certain attractive properties of updating. At first glance, this is bad news for the accuracy

[2]As well as theories of updating, some other evidential requirements that have come in for an accuracy-flavoured treatment are the principle of indifference and the principal principle (Pettigrew, 2016a), and other kinds of deference principles (Levinstein, in press; Dorst et al., 2021).

programme. With that said, I think there may be a way of understanding the problem so that it does not afflict the accuracy programme in particular, but rather arises out of more general tensions between fairly weak constraints on a theory of updating, on the one hand, and an attractive conception of evidence, on the other. If that is true, the problem may be one that everyone has to share, and not one that is particular trouble for the accuracy epistemologist. In that case, the accuracy epistemologist's hegemonic ambitions can remain intact.

Here is a guide to our discussion. In section 5.2 I comment on the externalist nature of these globally accuracy-conducive update functions. In section 5.3, I present three unusual features of these alternative ways of updating, while in section 5.4, I argue those features are not just unusual: they're outright undesirable. In section 5.5 I look at some fairly narrow ways of playing down the problems just identified, but suggest that none are particularly attractive. Then, in section 5.6, I argue that the accuracy epistemologist should instead claim that the problem I have identified is a much more general one, and that this partners-in-guilt style response absolves them of any particular responsibility to solve it.

5.2 Accurate updating and evidence externalism

In this section we discuss accurate updating in the context of both internalist, but more importantly, externalist theories of evidence. We see different ways of updating that plausibly maximise expected accuracy, depending on how exactly that latter notion is measured.

As we have seen already, the orthodox theory of updating — conditionalisation — says that upon receiving the evidence that p, your posterior probabilities should be your prior probabilities, conditional on p. And as we've seen, once we restrict our attention to the global way of evaluating update functions, as we are doing in

this chapter, conditioning only maximises expected accuracy if the set of propositions $\mathcal{E}$ you may receive as evidence form a partition. In the case that $\mathcal{E}$ doesn't, it may be more accuracy-conducive to update in another way. As we saw in the previous chapters, depending on exactly how we measure the accuracy of update functions, it might be best to update by meta-conditioning, update-conditioning or externalist-conditioning.

Given how we introduced internalism and externalism in the previous chapter, there's a 'stipulative' sense in which these these functions are 'externalist'; they all arise out of cases where you thinks your evidence is not perfectly informative. It is less clear that they are externalist in a slightly broader sense of 'externalist'. In particular, and as we will see, they can all require you to become certain of facts about your evidence, or your psychology, that externalists in this broader sense should think one cannot be rationally certain of. I think there are good reasons to worry that these update functions fail to meet the spirit of externalism, even if they meet the letter, and I will try to briefly note these as I go along. A little more directly, we might well worry that the informal philosophical motivations behind our stipulated meaning of 'externalism' (for example, that our evidence must be safe or known) stand in tension with some of the requirements these update functions will turn out to promote. Indeed, Miriam Schoenfield makes exactly this observation of meta-conditioning (Schoenfield, 2017, pp. 1180-86).

Less opaquely, it is notable that these update functions are extremely sensitive to features of your psychology, in a way that externalists in the broader sense might find questionable. As we've seen, update-conditionalisation requires you to become certain in facts about your psychological states even in the case where you have received evidence totally unrelated to your psychological states. Similarly, externalist-conditionalisation can have you become much more confident in a fact about your psychology even when your evidence was not about your psychology.

For those externalists who cleave a sharp line between your evidence and your (non-factive) psychological states, this is liable to look somewhat suspect: a form of internalism in sheep's clothing. Accordingly, and although it will not be the focus of our discussion in what follows, it's worth remembering that externalists in the broader sense may deny these are genuinely externalist update functions (and may reject the 'externalist-conditionalisation' label).

All three functions, whether interpreted as update rules or update dispositions, differ. We can see this by borrowing a helpful example from Gallow (2021): a simplified version of the unmarked clock case we met in the previous chapter, where there are only four positions of interest: 12, 3, 6 and 9. Suppose that at each position, for all your total evidence says, the hand is at the position either side (but not at the position opposite). Suppose moreover that, whatever total evidence you receive, there is a good chance your dispositions will react correctly, and a small chance your dispositions will react as though you are at one of the adjacent hands (but never as though you are at the position opposite). In Tables 5.1-5.5, taken and adapted from the very similar set of tables in Gallow (2021), all four update functions start with the same prior but yield different posterior probabilities when given the total evidence $\neg 9$ (that is, the total evidence you receive when the clock hand points at 3).[3]

To summarise, once we allow externalism about evidence, then given the global way of evaluating updates conditioning does not maximise expected accuracy — as we saw in the previous chapter. Which update function *does* max-

[3]The update functions do not disagree in every case. As mentioned earlier, when E is an equivalence relation then to meta-condition is just to condition. Moreover, meta-conditioning coincides with update-conditioning and externalist-conditioning in the case where $\mathcal{U} = \mathcal{T}$. The case for the coincidence of meta-conditioning and update-conditioning is straightforward; to see that meta-conditioning and externalist-conditioning coincide when $\mathcal{U} = \mathcal{T}$, note that for all $T(p)$, and for all $U(p)$, either $Pr(T(p) \mid U(p)) = 1$ or $Pr(T(p) \mid U(p)) = 0$. In which case, the updated probability reduces to $Pr(\cdot \mid T(p))$ for that $T(p)$ where $Pr(T(p) \mid U(p)) = 1$, which is just the meta-conditioning function.

	$12 \wedge T(\neg 6)$	$3 \wedge T(\neg 9)$	$6 \wedge T(\neg 12)$	$9 \wedge T(\neg 3)$
$U(\neg 6)$	$^{8}/_{40}$	$^{1}/_{40}$	$\mathbf{0}$	$^{1}/_{40}$
$U(\neg 9)$	$^{1}/_{40}$	$^{8}/_{40}$	$^{1}/_{40}$	$\mathbf{0}$
$U(\neg 12)$	$\mathbf{0}$	$^{1}/_{40}$	$^{8}/_{40}$	$^{1}/_{40}$
$U(\neg 3)$	$^{1}/_{40}$	$\mathbf{0}$	$^{1}/_{40}$	$^{8}/_{40}$

Table 5.1: Prior probability.

	$12 \wedge T(\neg 6)$	$3 \wedge T(\neg 9)$	$6 \wedge T(\neg 12)$	$9 \wedge T(\neg 3)$
$U(\neg 6)$	$^{8}/_{30}$	$^{1}/_{30}$	$\mathbf{0}$	0
$U(\neg 9)$	$^{1}/_{30}$	$^{8}/_{30}$	$^{1}/_{30}$	$\mathbf{0}$
$U(\neg 12)$	$\mathbf{0}$	$^{1}/_{30}$	$^{8}/_{30}$	0
$U(\neg 3)$	$^{1}/_{30}$	$\mathbf{0}$	$^{1}/_{30}$	0

Table 5.2: Updated with conditionalisation.

	$12 \wedge T(\neg 6)$	$3 \wedge T(\neg 9)$	$6 \wedge T(\neg 12)$	$9 \wedge T(\neg 3)$
$U(\neg 6)$	0	$^{1}/_{10}$	$\mathbf{0}$	0
$U(\neg 9)$	0	$^{8}/_{10}$	0	$\mathbf{0}$
$U(\neg 12)$	$\mathbf{0}$	$^{1}/_{10}$	0	0
$U(\neg 3)$	0	$\mathbf{0}$	0	0

Table 5.3: Updated with meta-conditionalisation.

	$12 \wedge T(\neg 6)$	$3 \wedge T(\neg 9)$	$6 \wedge T(\neg 12)$	$9 \wedge T(\neg 3)$
$U(\neg 6)$	0	0	$\mathbf{0}$	0
$U(\neg 9)$	$^{1}/_{10}$	$^{8}/_{10}$	$^{1}/_{10}$	$\mathbf{0}$
$U(\neg 12)$	$\mathbf{0}$	0	0	0
$U(\neg 3)$	0	$\mathbf{0}$	0	0

Table 5.4: Updated with update-conditionalisation.

	$12 \wedge T(\neg 6)$	$3 \wedge T(\neg 9)$	$6 \wedge T(\neg 12)$	$9 \wedge T(\neg 3)$
$U(\neg 6)$	$^{8}/_{100}$	$^{8}/_{100}$	$\mathbf{0}$	0
$U(\neg 9)$	$^{1}/_{100}$	$^{64}/_{100}$	$^{1}/_{100}$	$\mathbf{0}$
$U(\neg 12)$	$\mathbf{0}$	$^{8}/_{100}$	$^{8}/_{100}$	0
$U(\neg 3)$	$^{1}/_{100}$	$\mathbf{0}$	$^{1}/_{100}$	0

Table 5.5: Updated with externalist-conditionalisation.

imise expected accuracy depends on how exactly the expected accuracy of those functions is measured.

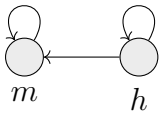
5.3 Three lessons about accurate updating given externalism

In this section, I want to provide three lessons about the accurate updating functions we met above.

Lesson 1

There is no asymmetry between the good case and the bad case.

Let's revisit the good case bad case scenario we saw earlier:



In the good case, you get useful information: you are in the good case! In the bad case, you get no information. The evidence you receive — W — does not rule out any worlds, and so, in an intuitive sense, does not allow you to cut down the space of possibilities. For all your evidence tells you, you could be anywhere.

That intuitive sense is well-reflected by conditionalisation: in the good case, you learn something, ruling out the bad case possibility. But in the bad case, where you learn nothing, your probabilities do not change. After all, your evidence gives you nothing with which you can rule out any worlds.

The first lesson about accurate updating given evidence externalism is that this asymmetry between the good and the bad case is illusory. Given that you learn

you are in the good case when you are, then even in the bad case, when you receive no evidence, you will still change your mind.

We can see this first with meta-conditionalisation, as noted by Das (2022, pp. 116-18). Since meta-conditionalisation requires you to condition on the proposition *that your total evidence is* such-and-such, rather than on the evidence itself, in the good case you again become certain you are in the good case, while in the bad case you become certain that you are in the bad case.[4] This is because the only world in which your total evidence is trivial is the bad case. Accordingly, the proposition that your total evidence is trivial just is the proposition that you are in the bad case. To become certain that your total evidence is trivial just is to become certain that you are in the bad case.

Importantly though, this isn't just a problem for meta-conditioning. Even if we adjust how the expected accuracy of the updates is being measured in order to vindicate a different rule, we end up with the same problem. First, update-conditionalisation. In the good case, your dispositions react as though you have

[4]A similar kind of being-required-to-go-beyond-your-evidence happens in unmarked clock cases: as Gallow (2021) points out, meta-conditioning will have you become certain you're at w_{52} when you are, in the example frame given earlier. This is another example of a case where an update function that is externalist in the stipulated sense given earlier nonetheless violates a broader externalist spirit; given any sort of reliability or safety condition on your beliefs, meta-conditioning in unmarked clock cases will lead one astray. That in this respect meta-conditioning is internalist-in-spirit was already noted by Schoenfield (2017) albeit not for unmarked clock cases. Specifically, Schoenfield noted that given meta-conditioning there is a set of propositions one can become certain of iff they are true, namely, the set $\mathcal{T}$. Obviously, this kind of 'luminous infallibility' sits uncomfortably with many externalist (in the broader sense) commitments. Crucially for Gallow's purposes, update-conditioning and externalist-conditioning *don't* have the property that in unmarked clock-style cases you can become certain of what evidence you have. And although update-conditioning also requires a set of propositions you will be certain in iff they are true (namely, $\mathcal{U}$), externalist-conditioning does not. Crucially for our purposes, the update functions we've been considering *don't* have to make a mess of non-transitive cases across the board. But they *do* have to make a mess of non-symmetric cases across the board, or so I argue above. Later I suggest that this indicates a tension between externalism and the accuracy framework. As we have just seen, we had some warning of this from Schoenfield (2017). But an interesting lesson of this chapter, I claim, is that even reconceiving how we measure the expected accuracy of update functions quite dramatically, and even defending update functions that avoid the problem Schoenfield was concerned with, is not by itself sufficient to get around the problem. The tension runs deeper than had previously been realised.

received the evidence that you are in the good case. Update-conditioning therefore amounts to becoming certain you are in the good case. In the bad case, your dispositions react as though you have received the evidence that you are in the bad case. Update-conditioning in the bad case therefore amounts to becoming certain you are in the bad case.

Next, externalist-conditionalisation. Suppose that, given you've updated on the evidence received in the bad case, you are antecedently pretty sure that you received the total evidence you would in the bad case (i.e. you received no evidence). This is enough to show that externalist-conditionalisation can require you to become more confident you are in the bad case when you are. As Gallow's externalist-conditionalisation update function has it, your probability you are in the bad case, given that you are in the bad case, is given by:

$$f(Pr, h)(\{h\}) = \sum_{q \subseteq W} Pr(\{h\} \mid T(q)) Pr(T(q) \mid U(\{h\}))$$

Since $T(\{m\})$ and $T(\{h\})$ partition W, this is equivalent to

$$Pr(\{h\} \mid \{h\}) Pr(\{h\} \mid U(\{h\})) + Pr(\{h\} \mid \{m\}) Pr(\{m\} \mid U(\{h\}))$$

which is equivalent to

$$Pr(\{h\} \mid U(\{h\}))$$

Assuming you are antecedently pretty sure you are in the good case, then $Pr(\{h\} \mid U(\{h\}))$ will be greater than $Pr(\{h\})$: in the bad case, you will become more confident you are in the bad case, despite receiving only trivial evidence.

There's a general structural reason for why these expectedly-accurate update

functions collapse the externalist asymmetry between the good case and the bad case. They all satisfy the following principle:

> **Reflection**[5]**:** Your prior probability in p should be identical to the expected value of your updated probability in p, that is, it should be that:
>
> $$Pr(p) = \sum_{w \in W} f(Pr, w)(p) \cdot Pr(U(E(w)))$$

What this says is that your present probabilities should be the same as the expectation of your future probabilities, post-learning-experience.[6] Given externalism, conditionalisation does not satisfy Reflection (Williamson, 2000; Salow, 2018; Das, 2022)). In good/bad-case scenarios like the one we have been considering, your probability that you're in the good case increases just in case you are in the good case, and stays the same otherwise. Since you think it's possible you're in the good case, you expect your probability that you're in the good case to increase. That is a violation of Reflection.

By the same token, in good/bad-case scenarios, update functions that satisfy Reflection and allow that your probability you're in the good case will increase when you are, must also allow that your probability you're in the good case must go down when you're in the bad case. That is, as a condition of Reflection and

[5]This exact formulation follows Gallow (2021, p. 495); cf. Salow (2018). See van Fraassen (1984) for very famous discussion of principles like this.

[6]See Gallow (2021, p. 495) for the explanation of why update-conditionalisation and externalist-conditionalisation satisfy Reflection. To see that meta-conditionalisation satisfies Reflection, observe that by the law of total probability and the fact that $\mathcal{T}$ partitions W:

$$Pr(p) = \sum_{q \subseteq W} Pr(p \mid T(q)) \cdot Pr(T(q))$$

Which is identical to

$$\sum_{w \in W} f(Pr, w)(p) \cdot Pr(T(E(w))) = Pr(p)$$

when f is the meta-conditionalisation function.

receiving useful information in the good case, these rules must allow that the probability you are in the bad case increases when you are in the bad case (see Salow (2018, pp. 702-707) for more on the relationship between Reflection and good/bad cases, and Das (2022, pp. 116-17) for discussion of this sceptical consequence specifically as it bears on meta-conditioning).

If you become more confident you are in the good case when you are, then given Reflection, you become more confident you are in the bad case when you are. If you wish to block the latter, this has the consequence that you cannot become more confident you are in the good case when you are: if you'll become no more or less confident you're in the good case when you're not, then you must also become no more confident you're in the good case when you are. As Das points out in a very slightly different context, the natural way of defending this view is by claiming that we can rule out sceptical hypotheses *a priori* — a controversial view, to put it mildly (Das, in press, pp. 27–28; see also Das (in press, p. 29) for this sceptical worry).[7]

In either case, then, these update functions must collapse the evidential asymmetry of the good and bad case. Either both worlds require belief change, or neither does. This concludes our first lesson.

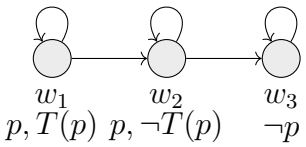
[7]White (2006) endorses this view. In fact, the good-case/bad-case scenarios Gallow discusses (see Gallow, 2021, p. 504-505) have this structure, though with a focus on the implications for defeat, rather than what the update rule means for the asymmetry. The models he looks at are more complex than the simple models we've discussed above, and as described above, they block your becoming more confident you are the bad case when you are by also blocking your becoming more confident you are in the good case when you are. Whether or not these models are realistic, they do nothing by themselves to undermine the conclusions we drew from working with the simple models above; that if you can become more confident you are in the good case when you are, it is at the cost of becoming more confident you are in the bad case when you are, too. With all this said, I'll confess to finding Gallow's example model here perplexing. In general, externalist-conditionalisation says you get your updated opinion in r given evidence p by taking the proposition $U(p)$ (the proposition that you've taken your evidence to be p and updated on it), then summing $Pr(r \mid T(q))Pr(T(q) \mid U(p))$ for every $T(q) \in \mathcal{T}$. So when your total evidence is W, one first needs to fix $U(W)$ to find the updated probabilities. But $U(W)$ is not even defined in Gallow's example.

Lesson 2

Updating is commutative, but not cumulative.

An update function is commutative iff receiving one proposition p as evidence, updating, then receiving another proposition q as evidence, and updating, outputs the same probability function as first receiving q, updating then receiving p and updating. In a slogan: order of evidence does not matter. Intuitively, commutativity is an attractive property of an update function. If I tell you that the murder victim was in the drawing room, and then next that they were killed with the candlestick, you should be in the same epistemic state as if I'd told you those things in the reverse order.

It's well-known that conditioning is a commutative update function. Whether the accurate update functions are commutative is slightly less clear. It's easiest to see why with meta-conditioning. Imagine you have a regular prior and the space of possibilities is encoded in the following frame.



Then suppose you are at w_1. So the evidence you receive is p. But since you update by meta-conditioning, you instead condition on $T(p)$. Whereas conditioning would amount to becoming certain that you are at either w_1 or w_2, meta-conditioning thereby amounts to becoming certain you are at w_1. Having meta-conditioned, we are left with the following frame:

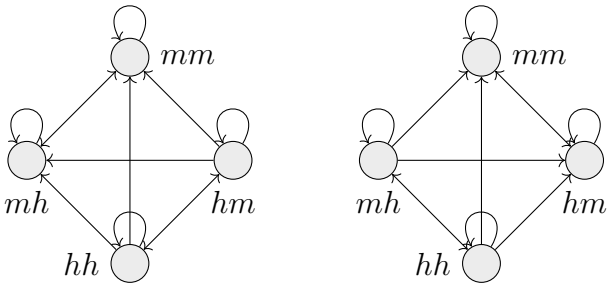


At this point, no further updates are possible. The only evidence you can receive is p, and updating again on $T(p)$ just returns the same frame.

More generally, recall that $\mathcal{T} = \{T(p) : \exists w : E(w) = p\}$, where $T(p) = \{w : E(w) = p\}$. Since the first update takes one to the set of worlds $T(p) \in \mathcal{T}$, for some p, then at every world in that set $T(p)$ is the case. So at every world in that set, the strongest evidence you can receive is p. No further updates are possible. Since update-conditioning exploits the same partitional structure as meta-conditioning, a similar problem will arise here, too (we will discuss externalist-conditioning in more detail later). Given the foregoing, the update functions are commutative, but trivially so. That is better than their being trivially non-commutative, but not by much.

We can ask more substantive questions about commutativity by slightly enriching our framework.[8] Return to our earlier case involving the fridge. Imagine now that there is not one fridge, but two, and each fridge either contains milk, or visually indistinguishable hoax milk. You will look in the left fridge, and then the right fridge; mm denotes that both of the fridges contain genuine milk, mh that only the left fridge does, and so on. Then we can capture your two looks into the respective fridges by associating a different accessibility relation to the respective looks, and intersecting your updated state at each. The first frame corresponds to looking in the left fridge, the second frame corresponds to looking in the right fridge.

[8] I don't claim this is the only way to handle this issue, but it's a fairly natural way of doing so.



Now suppose you first look in the left hand fridge, where there is hoax milk. Then you receive the trivial evidence W, and meta-conditioning on this, or equivalently, conditioning on $T(W) = \{hh, hm\}$ means you learn you are in one of these two worlds. Now you look in the right hand fridge; again, suppose you are at hh so there is hoax milk. Then meta-conditioning, by the lights of the second frame, or equivalently, conditioning on $T(W) = \{hh, mh\}$, means you learn you are in one of these two worlds. These two pieces of information combine to tell you that you are in hh. It's easy to see that order of information doesn't matter here, and that consequently, that you could have updated in either order to reach the same final state. More generally: since each update corresponds to locating yourself in a particular partition, the result is found by intersecting the two cells, and intersection is commutative, the results of the two updates will be, too.

With that said, certain other properties of updates are sacrificed. Here is a natural thought: if I give you some information, then once you've updated on that information, and assuming you have not changed your mind in the meantime, giving you that same information again should not change your mind.[9] For example, if I tell you that the butler did it, and a few seconds later, without having told you anything else, tell you again that the butler did it, then it's exceedingly natural to think that I've not given you any information that I hadn't given you before, and

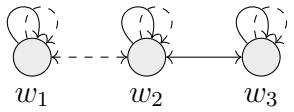
[9]Das (2022, p. 118) notes a version of this problem for meta-conditioning. If there is a standard name for this property, I'm not sure what it is. Das (2022, p. 118) calls it the 'fixity' principle. It also has some similarities to the weak preservation condition in the belief revision literature, see e.g. Shear and Fitelson (2019, p. 491). Thanks to Martin Smith for this suggestion.

that you shouldn't change your mind a second time. It's well known that conditioning respects this very natural thought: for all p, $Pr_p(\cdot \mid p) = Pr_p(\cdot)$. The accurate update rules do not. We can use the above example to show this. The evidence you receive in the first frame is the trivial evidence, W. The evidence you receive in the second frame is again the trivial evidence, W. So you have received the same evidence twice. Despite this, you continue to change your mind. After the first update, you know you are in either hh or hm (but not which). After the second, you know you are in hh. Since meta-conditioning is a special case of the other two update rules (see fn. 3) , those rules also fail to respect this property.

Another intuitively attractive property of updating is sometimes called 'cumulativity'. Intuitively, what it says is that if you receive p as evidence, then q, you should end up in the same place as if you had just received $p \cap q$. If I tell you the victim was murdered in the drawing room, and then I tell you they were murdered with the candlestick, intuitively you should learn just the same if I straightforwardly tell you that they were murdered in the drawing room with the candlestick. Again, conditioning respects this natural thought. And again, our accurate update functions do not. In fact, the previous example again suffices to show this. Since $W \cap W = W$, if the accurate update functions were cumulative then receiving W as evidence once should take you to the same place as receiving W as evidence twice. But as we've seen, that isn't what happens; receiving W the second time allows you to rule more worlds out. Again, since meta-conditioning is a special case of update-conditioning and externalist-conditioning (see fn. 3), those update functions have to share in the problem too. Nothing depends on it being the case that both updates involve receiving trivial evidence, either. Suppose you are at hm. Then after the first look, you receive trivial evidence, and by meta-conditioning, learn you are in the set of worlds $\{hh, hm\}$. After the second look, you receive the evidence $\{mm, hm\}$; meta-conditioning, you are in the set

$\{mm, hm\}$. Intersecting, you are at hm. The evidence you receive in the first update is W, in the second, $\{mm, hm\}$. Intersecting the two, you get $\{mm, hm\}$. But if you were just to receive this evidence, you would not learn you are at hm; in the second frame, meta-conditioning on $\{mm, hm\}$ tells you only that you are at $\{mm, hm\}$.

Since update-conditioning exploits the same partitional structure as meta-conditioning, those examples also cause trouble for update-conditioning. The situation with externalist-conditioning is slightly different. Consider the following frame



The lined arrows again denote evidential accessibility. A lined arrow from w to w' means that for all your total evidence says in w, you are in w'. A dashed arrow from w to w' represents what we might call 'update' accessibility. It says that for all your update dispositions respond to at w, they would respond the same way at w'.

Now suppose you start out with a regular, uniform prior. At w_1, after one update via externalist-conditioning, you will be $3/4$ sure that you are at $U(E(w_1))$. But after another update, you will be $5/6$ sure you are at $U(E(w_1))$. These again amount to violations of cumulativity; externalist-conditioning given evidence $E(w_1)$ followed by further externalist-conditioning given $E(w_1)$, results in a different probability function to the one given by externalist-conditioning given $E(w_1) \cap E(w_1)$.

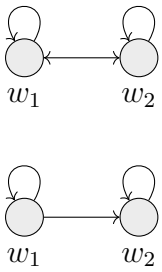
In sum, our lesson is that although accurate updating given evidence externalism is commutative, it is not cumulative.

Lesson 3

Updating is highly perspective-dependent.

Our final lesson about accurate updating given evidence externalism is this. It is highly perspective-dependent. Of course, there is a sense in which even orthodox conditioning is perspective-dependent, since it exploits your conditional probabilities, and those form part of your perspective. But these variant update functions are more perspective-dependent, still. Specifically, they do not react only to what evidence you receive, but what evidence you might have received (but did not). This means that two agents can have the same priors, and receive the same information, but end up with distinct posterior probabilities.[10]

Here is an example involving meta-conditioning. Let the first frame capture the situation of the first agent, and the second frame capture the situation of the second. Assume moreover that they have identical priors.



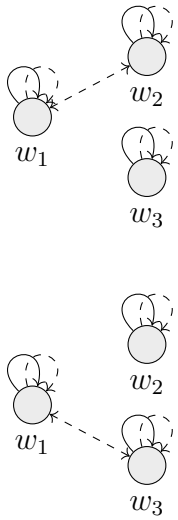
Suppose both agents are at w_1. So both receive the (trivial) evidence $\{w_1, w_2\}$. Despite this, upon meta-conditioning, the first agent will undergo no change in view; meta-conditioning on $\{w_1, w_2\}$ is just to condition on the trivial proposition

[10]As I discuss further in fn. 18, Bronfman (2014) already notices that deviating from a rule like meta-conditionalisation (as Gallow eventually does) might introduce perspective-dependence. Bronfman is interested in a strengthening of the availability condition; the informal constraint Gallow places on the set of update functions through $\mathcal{U}$ is in general neither a strengthening nor weakening of availability but still leads to perspective-dependence. Crucially though, and as we will see shortly, even meta-conditioning introduces a high degree of perspective-dependence.

$\{w_1, w_2\}$ given the nature of the frame. However the second agent will become certain they are at w_1, for w_1 is the only world in which they can received $\{w_1, w_2\}$ as evidence.

Even though the two agents have the same priors and receive the same information, they end up with distinct posterior probabilities. That is entirely because they have different views of what they might receive as evidence. Therefore, despite *actually* receiving the same propositions as evidence, they end up with different opinions after updating.

Since update-conditioning and externalist-conditioning collapse into meta-conditioning in the case where $\mathcal{T} = \mathcal{U}$, technically speaking the above example is enough to show that these update functions are also highly perspective-dependent. That said, it might be insisted that there are no cases in which those two sets coincide, in which case we will need to provide different examples.



Again, let the first frame capture the situation of the first agent, and the second frame capture the situation of the second. The lined arrows again denote evidential accessibility. A lined arrow from w to w' means that for all your total evidence says in w, you are in w'. A dashed arrow from w to w' again represents what

we might call 'update' accessibility. It says that for all your update dispositions respond to at w, they would respond the same way at w'. For ease, assume a) the agents have uniform regular priors (though the point survives relaxing uniformity at the cost of very slightly more complexity) and b) both agents are at w_1.

First we can see what happens in the case of update-conditioning. The first agent will condition on the proposition that they've updated as though they've received the proposition given as total evidence at w_1, which is to update the same way as though they'd been at w_2 (and receiving the evidence they would have received there). So the agent will become certain they are at w_1 or w_2. Whereas the second agent will condition on the proposition that they've updated as though they've received the proposition given as total evidence at w_1, which is to update the same way as though they'd been at w_3 (and receiving the evidence they would have received there). So the agent will become certain they are at w_1 or w_3. In sum, the agents will end up with different posterior probabilities, despite having the same prior probabilities and receiving the same evidence.

Similarly for externalist-conditioning. Recall that externalist-conditionalisation is formalised as follows:

$$f(Pr, w)(\cdot) = \sum_{q \subseteq W} Pr(\cdot \mid T(q)) Pr(T(q) \mid U(E(w)))$$

Because of the way $\mathcal{U}$ and $\mathcal{T}$ are related in the two frames, and because the agents have uniform regular priors, then in the first frame:

$$f(Pr, w_1)(\cdot) = 1/2(Pr(\cdot \mid T(E(w_1))) + Pr(\cdot \mid T(E(w_2))))$$

and in the second frame:

$$f(Pr, w_1)(\cdot) = 1/2(Pr(\cdot \mid T(E(w_1))) + Pr(\cdot \mid T(E(w_3))))$$

But then it is easy to find cases where the posterior probabilities disagree, despite identical priors and evidence. For instance, both updated probability functions will disagree over $T(E(w_1))$ and $T(E(w_2))$. So again, we have a case where your posterior probabilities do not simply depend on the evidence received and your prior opinions.

The lesson is that accurate updating given evidence externalism is highly perspective-dependent. These update functions are sensitive not only to what evidence you have, but what evidence you might have had.

5.4 The bad news

On the face of it, I think what we've learned is surprising. Accurate updating given evidence externalism has several properties we don't associate with the orthodox theory of updating — that is, conditionalisation. But more importantly, as I now want to claim, it lacks several properties that we typically think of as important good-making features of evidence-responsiveness. Those properties are possessed by conditionalisation, but crucially, they're also attractive independently of that specific update function.

First, the update functions collapse the asymmetry between the good and the bad case. In doing so, they make one of two errors. First, they may require that you become confident you are in the bad case, when you are. In doing so, they seem to impose beliefs which fail to meet certain epistemic standards; they are unsafe, unreliably formed, ungrounded in evidence. Second, they may require that you do not become confident you are in the good case, when you are. In doing so, they concede too much ground to the sceptic, who thinks you can never learn you are in the good case. Conditioning, however, preserves the good-bad case asymmetry. It allows that you can learn you are in the good case, while forbidding becoming

confident you are in the bad case.

Second, the expectedly accurate update functions are not cumulative. Given these functions, there is a difference between receiving p as evidence, then receiving q as evidence, and receiving $p \cap q$ as evidence. But intuitively, there is no such difference: a judgment conditionalisation respects.

Lastly, expectedly accurate updating forces your updating to depend not only on your prior probabilities and the evidence you *do* receive, but on the evidence you didn't receive, yet might have. Accordingly, agents with identical priors can receive identical evidence but end up with different beliefs, despite using the very same update function. Intuitively, the only thing on which your update should depend is your prior belief state, your update function and the evidence you receive: again, something conditionalisation manages to respect while the other functions do not.

Insofar as the expectedly-accurate update functions work in this way, the goal of capturing all evidential norms in accuracy-theoretic terms will falter. We take the above to be desirable features of learning independently of any commitment to conditionalisation. Rather, they just seem like features of responding to evidence that a theory of learning should capture. Insofar as these update functions fail to capture them, that is a problem for the accuracy programme's hegemonic ambitions.[11]

[11]It's interesting to compare Das (2022, p. 119); he notes (effectively) given the issues first identified by Schoenfield (2017), there will be problems for giving an accuracy-centred justification of conditionalisation. But what we've seen is that the situation is worse than that, because the accuracy-motivated replacements for conditioning — including those replacements that, per Gallow, are meant to be better than meta-conditioning — also have a host of very undesirable properties.

5.5 Some unpromising ways out

With this in mind, it's worth checking whether there are ways of playing down the strangeness of some of the above. We'll look at three possible responses in this section. One involves attempting to play down the badness of collapsing the good/bad-case asymmetry. Another involves reconceiving the 'evidence for' relation in light of the new update rules. The last attempts to ground the strange properties by understanding the variant update rules as parts of a non-idealised epistemology. None look particularly promising.

5.5.1 Why worry about asymmetry

The first thing I want to discuss is whether the collapse of the good/bad-case asymmetry can be ameliorated. To recall: what we learned was that either you become confident you are in the good case when you are, iff you become confident you are in the bad case when you are, or, you cannot become confident you are in either the good or bad case. It is worth noting that anti-scepticism is perhaps not especially motivating for everyone, in which case, it might be tolerable to hold on to the idea that you cannot become more or less sure you are in the good case, wherever you find yourself. For example, you might endorse an epistemological ideology wherein knowledge is de-prioritised. Consequently, the importance of being able to know that one has hands will be de-prioritised, too (as might an attendant 'becoming confident one has hands'). Indeed, as we described earlier, the value theory that accuracy epistemologists typically ascribe to themselves is a 'veritist' one (Pettigrew, 2016a), akin to the veritist view in 'traditional' epistemology wherein true beliefs are the fundamental bearers of epistemic goodness (Goldman, 1999). Alternatively, one might think it is important that one believes (is confident in) anti-sceptical hypotheses but hold that we have prior 'a priori'

confidence that we are in the good case, meaning that not being able to increase or decrease that confidence is tolerable (as we alluded to earlier). I don't want to get too mired in these issues, so I will just note a couple of things. First, at least some externalists *are* moved by these sorts of anti-sceptical considerations. In which case, it's interesting to find out that the expectedly accurate update functions stand in some tension with that philosophical outlook. Second, it is worth emphasising that the equation of non-symmetric frames with drastically sceptical scenarios like the ones we considered earlier is not guaranteed. Some are altogether more realistic. Here is an attractive example from Maria Lasonen-Aarnio.[12]

> I take the following to be a pretty common phenomenon: coming across a fake, one can mistake it for the real thing, but when one sees (feels, hears, tastes, smells) the real thing, one can tell that it is not a fake. This is one reason why verbal directions are bad. It is easy to get lost by taking the little, barely visible path, where the guidebook tells one to take the *obvious* path to the left, and not to get distracted by tiny, barely visible paths. When seeing one of the small paths, it is far from obvious that it is not obvious — after all, it is a path. Upon finally finding the right path, one immediately recognizes that this is what the guidebook was talking about. (Lasonen-Aarnio, 2015, p. 160)

This is a quotidian case. But plausibly it has the same asymmetric structure seen in more dramatic examples involving hoax milk or brains in vats. When faced with the obvious path, you knows you are faced with the obvious path. When faced with the little path, for all you know, you are faced with the obvious path, but for all you knows, you are not. The accurate update rules will all collapse

[12]Das (2019, p. 84) presents another realistic example of negative introspection failure.

this symmetry. If you update via meta-conditioning, then faced with the little path, you should become certain you are in the wrong place. Here's another example. A recent plumbing issue meant I had to search underneath the kitchen sink for the stopcock (several times, in fact). The valve is sufficiently stiff that it's not possible to tell by hand whether it needs to go clockwise for on, or off. Because of this, you need to check by sight which way it should be turned, and fortunately, the tap is labelled 'on' and 'off', with corresponding arrows to show which way it should be turned. Unfortunately, the space is very poorly lit, and it can be difficult to make out the words 'on' and 'off' in the following, interesting way. If you see the word 'on', it's unclear whether it is in fact the word 'on' or 'off' — you can make out the 'o', a character next to it, and nothing else. It's consistent with this that it says 'off'. If you see 'off', however, the two 'f's are visible in such a way that you can tell you are looking at 'off' and not 'on'. This gripping tale is another example, I would suggest, of the asymmetric structure we are interested in.

Cases like these make me somewhat doubtful that the collapse of asymmetry these update functions cause can be explained away, even by externalists not especially concerned about the issue of scepticism.[13]

5.5.2 Reconceiving the 'evidence for' relation

Another strategy for playing down the unattractive features of the alternative update functions is to claim they are less surprising or bad than they first seem because alternative update functions necessitate alternative conceptions of 'evidence for': alternative conceptions that can explain away some of this strangeness.

To expand on this: we've taken, as is standard, $Pr(p \mid q)$ to be the conditional probability of p given q, that is, $\frac{Pr(p \cap q)}{Pr(q)}$. We can also think of this as something

[13]Lasonen-Aarnio (2015) discusses other ordinary cases with this apparent structure, though see Salow (2018) for a challenge to the claim these cases in fact have this non-partitional structure after all.

like your future opinion in p upon learning q, *assuming* that conditioning is the right theory of learning. Obviously, we cannot assume that in our case. So it is worth considering alternative ways of understanding the notion.

In particular, let:

$$\overset{M}{Pr}(p \mid q), \overset{U}{Pr}(p \mid q) \text{ and } \overset{E}{Pr}(p \mid q)$$

stand for the posterior probability you have in p given you have received q as evidence and updated according to meta-conditionalisation, update-conditionalisation and externalist-conditionalisation, respectively. That is, they are identical to $f(Pr, w)(p)$ where f is, respectively, the meta-conditioning, update-conditioning and externalist-conditioning function, and $E(w) = q$. Consequently, they are only defined for $q \in \mathcal{E}$.[14] A popular (if not absolutely universal) formalisation of the 'evidence for' relation, in the context of standard conditioning, is the following.

Evidence for: p is evidence for q iff $Pr(q \mid p) > Pr(q)$

The idea is roughly that p is evidence for q if receiving p as evidence raises your probability in q. With our alternative theories of learning on the table, we can encode alternative 'evidence for relations', where:

$$\overset{M}{Pr}(p \mid q) > Pr(p);\ \overset{U}{Pr}(p \mid q) > Pr(p), \text{ and } \overset{E}{Pr}(p \mid q) > Pr(p)$$

Thinking of things in these terms definitely goes *some* way to undoing the awkwardness of some of the above lessons. For instance, it's on the face of it really surprising that getting trivial evidence can bring about a change in view, given these alternative update functions. It's maybe less surprising when we realise

[14] I mean this as a technical point, not an objection; the same thing is true for the conditioning update function, for instance.

that, unlike conditioning where for all p, $Pr(p \mid W) = Pr(p)$, for the other update rules there can be propositions q such that

$$Pr(q \overset{M/U/E}{\mid} W) > Pr(q)$$

In other words, given these alternative update functions, and corresponding notions of 'evidence for', the tautology *can* be evidence for other propositions: contrary to the usual treatment of the bad case, given conditioning.

That said, I wonder if this is really very satisfying. For one thing, it *just is* quite surprising that the tautology can be positively (or negatively) relevant to any proposition, independently of exactly how you feel about updating. While we can manipulate our concept of 'evidence for', it's another thing entirely to manipulate our concept of 'evidence', and as understood in the setting, trivial evidence is just that: it tells you nothing about where you are. For another thing, thinking of things in this alternative way does nothing to change some of our other lessons. Externalists should presumably still balk at the idea you can learn you are in the bad case when you are.[15] Maybe the troubling observation that you can learn you are in the bad case can be mitigated by appeal to a reformed notion of 'evidence for', but that one has evidence for this proposition in such a scenario ought to be very surprising. Likewise, externalists should presumably still balk at the idea you cannot learn you are in the good case when you are. Similarly, although you can now be understood as having evidence *for* different propositions in the pairwise agential comparisons of lesson three, it remains very surprising that what you have evidence for can vary so dramatically between agents who share the same evidence and the same priors.

There are other wrinkles with this response. Nothing in the characterisation

[15] See Magidor (2018) for dissent on this, though note her reasons for thinking you can learn you are in the bad case are structurally quite different, in effect amounting to a denial you receive no evidence in the bad case.

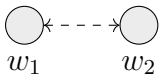
of U seems to require that $U(p)$ entails p. It doesn't even entail $T(p)$, for Gallow wants to allow that sometimes $Pr(T(p) \mid U(p)) < 1$; you can update/externalist-condition and remain uncertain what your total evidence is. But then there's nothing *obviously* in the characterisation of U to force the result that $Pr(p \mid U(p)) = 1$. This means that it's consistent with this alternative evidence-for relation that

$$Pr(p \overset{U}{\mid} p) < 1$$

And relatedly, that there can be cases where

$$Pr(p \overset{U}{\mid} p) < Pr(p)$$

as in the following example:



Let w_1 be the only world where p is true; w_2 is then the only world where one updates as though p is true. Assuming, for instance, a regular prior, then conditioning on $U(p)$ will reduce your confidence in p.

I think this is quite hard to swallow. Similar remarks will apply to externalist-conditioning.[16] *Perhaps* it's tolerable that you can receive a proposition as evidence, update accordingly but remain less than certain in it. But the idea that a proposition is not evidence for itself seems to stretch the concept of 'evidence for' to breaking point.[17]

[16] And in fact, an instance of $Pr(p \overset{E}{\mid} p) < 1$ appears in Gallow's example (Gallow, 2021, fig. 9), though obviously not in terms of an alternative evidence-for relation. Note that similar remarks *don't* apply to meta-conditioning because we assume evidence is factive; it's much less clear why you should think your update dispositions can never be so drastically misled.

[17] A little more carefully, I mean the technical, probabilistic conception of 'evidence for'. The 'probability raising' view mentioned above remains popular, though see Brown (2018, ch. 3) for some recent criticism.

There's a more general worry about supplementing usual conditional probabilities with this alternative notion. Specifically, it's unclear what the limits of this re-casting of the role of conditional probabilities are supposed to be. To repeat, one way of thinking about conditional probabilities is as your updated belief state given the proposition received as evidence, where the update function is conditioning. In that case, we should substitute it with the relevant, expected-accuracy-maximising update function as needed. This is particularly important because conditional probabilities figure in at least some of the measurements of the expected accuracy of update functions. The externalist-conditioning function is found to be the most expectedly-accurate update function when that expected accuracy is measured in the following way:

Definition 166.

$$EA^{\prime\prime c}_{Pr}(f) = \sum_{p \subseteq W} \sum_{w \in T(p)} Pr(\{w\}) \sum_{q \subseteq W} Pr(U(q) \mid T(p)) A^c(f(Pr, w), w))$$

As we described before, the informal gloss on this is supposed to be that: you think of all the pieces of information you might receive, then assess how accurate you are by measuring the accuracy of your update dispositions in response to that information, weighted by how likely it is those dispositions will 'fire' or 'misfire' (Gallow, 2021, p. 497). But presumably, if you endorse the externalist-conditionalisation update function, you should regard the following as your opinion of how likely you are to update correctly given particular evidence:

$$EA^{\prime\prime c}_{Pr}(f) = \sum_{p \subseteq W} \sum_{w \in T(p)} Pr(\{w\}) \sum_{q \subseteq W} Pr(U(q) \overset{E}{\mid} T(p)) A^c(f(Pr, w), w))$$

$Pr(U(q) \overset{E}{\mid} T(p))$ is only defined for $T(p) \in \mathcal{E}$, and of course there is no gen-

eral requirement that for all p, $T(p) \in \mathcal{E}$; quite the reverse if we are interested in cases of partitionality failure. This means that the above sum will not in general be defined, either, and it will not be possible to assess how the sum using the externalist-conditionalisation notion compares to the sum without. Consequently, it is not possible to say whether the expected-accuracy maximising dispositions are the same once we restate the initial sum to incorporate this revised approach to learning. This means there's an outstanding tension in the accuracy argument for those learning dispositions; they implicitly rely on a theory of updating that they then go on to reject.

In sum, it doesn't look like reimagining the 'evidence for' relation, or otherwise recasting conditional probabilities, in light of the various alternative theories of learning will be a promising way to explain away the surprising features of the accurate updating described above.

5.5.3 Appealing to non-idealised rationality

Perhaps another way to go is to claim that update-conditioning and externalist-conditioning, with their emphasis on highly contingent update dispositions, in fact provide update functions of a piece with *non-ideal* rationality. The story as I am envisioning it exploits the fact that we have started with the inevitable, human limits on the discriminative capacities of our dispositions — encoded in the partition $\mathcal{U}$ — and attempted to find the dispositions that will maximise expected accuracy in light of this. The result is update-conditioning or externalist-conditioning, depending on how exactly you think that expected accuracy should be measured. Seen in this light, we can again attempt to play down at least some of the strange results we met above. For example, the high degree of perspective-dependence may now look more defensible. Since we start with the discriminative limits on our update dispositions, and those may vary from person to person, perhaps it is

not so bad if the result of updating does so, too. The story would be that what we've learned, in effect, is that epistemic rationality is interpersonally permissive: two agents can share all the same evidence, but rationally react differently to some new evidence.[18] More generally, on this way of setting things out, the lesson is just that non-ideal rationality has all of the features described in section 5.3.

The version of permissivism one gets out of these update functions may be acceptable. It's worth noting that this kind of permissivism is slightly different to the version of permissivism that sees interpersonally divergent responses to new evidence as consisting in interpersonally divergent epistemic standards. For example, someone with 'risk-taking' standards might become much more confident upon receiving some evidence while someone with 'risk-averse' standards might become only slightly more confident given that same evidence.[19] Unfortunately, that motivation for permissivism does not latch very neatly onto the kind of permissivism you get out of these update functions. At best, the permissivism is grounded in differences in your update dispositions. Of course, those differences *might* reflect the two agents' differing standards. But there is nothing in the models to require this, and even if two agents had identical epistemic standards, it's consistent with the update functions that they could be interpersonally permissive nonetheless. Instead, and in the spirit of the response we are considering, it's best to try and locate the source of the permissivism in those differing dispositions

[18] As mentioned above, Bronfman (2014, pp. 887-88) notes that availability may be too weak to capture evidence-dependence, and suggests a further constraint; in doing so he anticipates that the result of modifying availability — as Gallow in effect does — may be that one ends up with the sort of perspective-dependence we have seen; "The precise determination of the available update rules may be difficult; it will depend on what the agent's abilities are" (Bronfman, 2014, p. 888). Schoenfield (2017, p. 1177) makes a similar observation about Bronfman's proposal: "[i]f we want to account for the limitations of non-ideal agents this [agent-relativity] is to be expected, but we are now quite far from the project as Greaves and Wallace, and others involved in accuracy-first epistemology, originally conceived of it" (Schoenfield, 2017, pp. 1177-78). We'll see above in considering the response from non-ideal rationality that in fact, we will end up with something agent-relative even given meta-conditioning, a highly idealised conception of epistemic rationality, and no strengthening or modification of availability whatsoever.

[19] See e.g. Kelly (2013); Schoenfield (2014), as we discussed in earlier chapters.

themselves. The idea — at least insofar as the permissivism arises from update-conditioning or externalist-conditioning — would have to be that the demands of non-ideal rationality are very closely tied to your dispositions, how good they are at recognising the total evidence you have received, and so forth.

With this said, I am still sceptical that the appeal to non-ideal rationality can solve all the accurate externalist's problems. For one thing, it does little to explain why the accuracy maximising updates are non-cumulative. There's not anything *obviously* in non-ideal rationality to explain this away. For another thing, on this way of looking at things, the non-idealising happens by letting the set $\mathcal{U}$ be distinct from the set $\mathcal{T}$. This is what allows two individuals to update in different ways despite receiving the same total evidence. In that case, the attendant *idealised* conception of updating is just meta-conditioning; that is what we get when we require the identity of $\mathcal{U}$ and $\mathcal{T}$. Consequently, idealised rationality also involves a high degree of perspective-dependence, struggles to make sense of anti-scepticism, and non-cumulative updating; all things we tried to explain away as features of non-ideal rationality. It's therefore unlikely this strategy can explain away these properties. Relatedly, the picture of ideal rationality this strategy would commit us to also looks unattractive. Indeed, once we realise these things, the identification of the source of the permissivism as differing dispositions looks a bit suspect; even the analogue idealised function, meta-conditionalisation, is permissive, despite making no reference to dispositions whatsoever.[20] And again, it's a bit unclear how that permissivism can be explained by appeal to one of the more natural stories, like differing epistemic standards. Nothing in the examples that show meta-conditioning to be highly perspective-dependent force this interpretation, making any such explanation fragile.

[20]Even if one buys the argument from non-ideal rationality, it still remains that our inquiry has unearthed the surprising fact that non-ideal rationality lacks a number of the evidential virtues we thought it had.

5.6 A more promising way out

We've looked at three features of accurate updating given evidence externalism. What we've seen is that there is quite a deep tension between certain evidentialist requirements associated with updating, the accuracy programme, and externalism about evidence. Certain ways out of this tension, described above, do not look like very promising ways to remove it.

On the face of it, this is a bad problem for the accuracy programme and its hegemonic ambitions. It looks as though certain epistemic virtues — in particular, certain evidential virtues — cannot be explained in terms of accuracy. With that said, I think there may be a way for the accuracy epistemologist to maintain their position. The way to do so is to deny that the problem described here is one that they alone face.

First, we should note that strict propriety is not obviously an accuracy-theoretic notion, in the way that, e.g. truth-directedness, which we met in previous chapters, is:

Definition 16 (Truth-directedness)**.** *If Cr and Cr' are credence functions, and $Cr(p) \geq Cr'(p)$ for all true propositions at w, $Cr(p) \leq Cr'(p)$ for all false propositions at w, and either $Cr(p) > Cr'(p)$ for some true proposition at w or $Cr(p) < Cr'(p)$ for some false proposition at w then $A^c(Cr, w) > A^c(Cr', w)$.*

It's plausible to think even non-accuracy epistemologist epistemic utility theorists should endorse strict propriety, as we argued in chapter one. This is already enough to identify a fair few partners in guilt for the accuracy epistemologist: anyone working in a broadly epistemic utility-theoretic framework. In fact, it may be possible to identify yet more. Even those working in non-consequentialist settings are open to being 'consequentialised'.[21] For some theory to be conse-

[21] See e.g. Portmore (2007); Horowitz (2017, pp. 286-87) discusses the prospects for interpret-

quentialised is for a version of that theory to be given which is extensionally equivalent to their theory — that is, agrees with theory in all obligations and permissions — but is formally consequentialist. Typically this is accomplished by invoking a gerrymandered utility function which guarantees that all those actions required by the non-consequentialist theory maximise expected utility, and all those which are prohibited do not. Given the general theoretical sympathy to (intrapersonal) uniqueness-type constraints, it is quite plausible that the utility function of the consequentialised theory will also have to satisfy strict propriety. Even non-consequentialist theories will want to capture this self-recommending property.[22]

All this is to say that, it is likely that a great many theorists are in effect committed to strict propriety. Since strict propriety was the key assumption needed to generate the arguments for these variant-conditioning functions, it is possible many theorists can be forced into this tension: evidence externalism, a natural self-recommendation feature of your view of epistemic requirements, and a natural way of assessing update functions all combine to show that certain apparent evidential virtues do not hold, after all. Consequently, it's not completely clear there is anything here that should concern accuracy epistemologists *in particular*. Instead, there may be a more general tension between strict propriety on the one hand, and evidence externalism on the other.

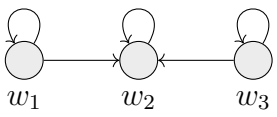
I think it is not at all clear how this tension should be resolved. On the one hand, we have very good theoretical reasons to endorse evidence externalism, many of which we described earlier. For many epistemologists, one of the key lessons of recent epistemology just is that evidence is not in general partitional.

ing accuracy epistemology as a consequentialised version of some non-consequentialist epistemic theory

[22]See e.g. D. Lewis (1971); Elga (2007) for theories which endorse the immodesty idea often taken to underly strict propriety despite being, if not explicitly non-consequentialist, then certainly not explicitly consequentialist.

On the other hand, strict propriety is extremely attractive. I don't know of any really telling objections to it, and the most promising ones instead enforce a move to propriety — given which, the arguments of this chapter go through virtually unchanged: instead of being required to update in an irrational way, you'll only be permitted to do so, which is hardly much of an improvement. Moreover, as we saw in chapter one, much of the weight of the accuracy programme rests on propriety's shoulders. The costs of rejecting it are severe.

A natural strategy at this point is to try to assimilate the problems we have encountered to other, similar, problems that evidence externalists must face. Specifically, it is helpful to recall a somewhat similar challenge discussed by Timothy Williamson in relation to non-partitional frames. He notes that such frames can lead to violations of the Reflection principle we discussed earlier. Specifically, there can be cases where you expects that your probabilities post-evidence will be different to your present probabilities (Williamson, 2000, pp. 230-37; Salow, 2018.)



Assume a regular uniform prior. Then this frame leads to violations of Reflection. For instance, your prior probability in $\{w_2\}$ is $\frac{1}{3}$. But if you are at w_1 or w_3 it will increase to $\frac{1}{2}$; if you are at w_2 it will increase to 1. So you expect your probability in $\{w_2\}$ to increase post learning-experience to at least $\frac{1}{2}$ regardless of what evidence you receive.

Given that you know this, it's at least somewhat tempting to think you should increase your probability to at least $\frac{1}{2}$ already, even in advance of receiving the evidence. After all, you know you will end up doing this anyway. As Williamson notes, the point can be made more 'vivid' by imagining that before any update,

you are offered a bet on the proposition $\{w_1, w_3\}$. The bet pays \$6 if $\{w_1, w_3\}$ is true and nothing otherwise. Given your uniform regular prior, the expected value of the bet is \$4, and you should be willing to pay up to \$4 for it. Yet on your posterior probabilities, you will be at most $\frac{1}{2}$ confident in the proposition (at w_1 and w_3) and possibly even certain of its falsity (at w_2). So at any world, you should accept any price of \$3 (or higher) for the bet. A 'shark' can guarantee themselves a profit by selling you the ticket before your update and buying it back after. Yet at each stage, the transactions maximise expected utility by your lights. Though you know in advance you will be exploited, at no point at which you are actually being exploited do you know you as much. Suppose you are at w_1. Then post update, your probability in $\{w_1, w_3\}$ is $\frac{1}{2}$. But it's not certain on your evidence that it's a half, because if it was, you would know you were at w_1, contrary to supposition.

Williamson's claim is that we have to live with the failures of the reflection principle, and the corresponding exploitation, because we cannot make use of information in advance of actually receiving it: "we cannot take advantage of the new knowledge[, conditioning on which would lead to greater probability in $\{w_2\}$,] in advance. We must cross that bridge when we come to it" (Williamson, 2000, p. 237).

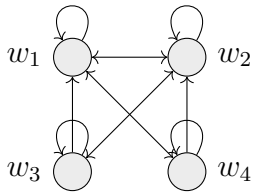
If this kind of 'exploitability' comes with the territory of evidence externalism, it's natural to hope that the problems we have encountered are just another kind of this issue. And at least superficially, I think there is a case for this. Compare, for instance, meta-conditioning. Meta-conditioning is better for us, in the sense of the expected values of the update functions: we expect it to make us more accurate than conditioning. But, as we discussed above, it requires us to take advantage of information we don't have. In meta-conditioning, we will not in general know the propositions that we are conditioning on; we are expected to make use of information in advance of receiving it; in fact, often to make use of information

we will never receive at all. In effect, this should be what externalists object to: the planning framework, as it has been used, treats as available plans that are *not*, because said plans require us to have access to information we don't, in fact, have access to. As we saw when considering the good case/bad case asymmetry, similar remarks will apply to update-conditioning and externalist-conditioning, too.

While I think this way of resolving the tension is initially attractive, I have (at least) a couple of worries. It is plausibly a prediction of the previous paragraph that, were we able to more carefully constrain the set of available update functions, conditioning would emerge as the update function which maximises expected accuracy. However there are (at least) two challenges with imposing further restrictions on the set of available update functions. First, it may be difficult to restrict the set of functions in such a way that any resulting accuracy argument is dialectically effective. To give an extreme example, if we constrained the set of available update functions in such a way that conditioning was the only available update function, then we would have a) an argument in this framework that conditioning maximises expected accuracy and b) persuaded absolutely no-one who was not already convinced of this.

Second, insofar as there have been attempts to more carefully constrain the set of available functions, they suggest that conditioning is the not the function that maximises expected accuracy. Specifically, Bronfman (2014, p. 886-90) gives a toy example wherein a further constraint is placed upon the set of update functions that both excludes meta-conditioning and does not exclude conditioning. Nonetheless, in his example conditioning fails to maximise expected accuracy.[23] Here is the frame Bronfman considers.

[23]Bronfman measures accuracy using a standard quadratic accuracy measure. I do not think rejecting this particular measure would be a fruitful way of rebutting his point, or the other points that arise out of his example.



The prior probabilities are $Pr(\{w_1\}) = Pr(\{w_2\}) = \frac{1}{6}$, and $Pr(\{w_3\}) = Pr(\{w_4\}) = \frac{1}{3}$. Here are some important observations about the frame. It is not partitional, for instance because transitivity fails: Ew_2w_1, Ew_1w_4, but not Ew_2w_4. Also, $T(\{w_1, w_2, w_4\}) = \{w_1, w_4\}$ while $T(\{w_1, w_2, w_3\}) = \{w_2, w_3\}$. Bronfman preserves the availability constraint we met earlier, that if $E(w) = E(w')$, $f(Pr, w) = f(Pr, w')$. He imposes a further constraint: that upon your total evidence being $\{w_1, w_2, w_3\}$, you are not more or less certain, post-update, of $\{w_1\} \cap \{w_1, w_2, w_3\} = \{w_1\}$ than $\{w_2\} \cap \{w_1, w_2, w_3\} = \{w_2\}$ — and similarly for your total evidence being $\{w_1, w_2, w_4\}$. The motivation for this is that your evidence does not tell you what your total evidence is in the frame.

Crucially, Bronfman's restriction rules out meta-conditioning as an available rule. If you were at w_1, then meta-conditioning would lead you to become certain they were at $\{w_1, w_4\}$. But then you would be more confident of $\{w_1\}$ than $\{w_2\}$ post-update, contrary to Bronfman's second restriction. Nonetheless, as Bronfman points out, conditioning is not the update function that maximises expected accuracy. Given the conditioning update function f, then $f(Pr, w_1) = f(Pr, w_4) = Pr'$, where $Pr'(\{w_1\}) = Pr'(\{w_2\}) = \frac{1}{4}$, $Pr'(\{w_3\}) = 0, Pr'(\{w_4\}) = \frac{1}{2}$; and $f(Pr, w_2) = f(Pr, w_3) = Pr''$, where $Pr''(\{w_1\}) = Pr''(\{w_2\}) = \frac{1}{4}, Pr''(\{w_3\}) = \frac{1}{2}$ and $Pr''(\{w_4\}) = 0$.

Yet Bronfman shows that given the update function f' that in fact maximises expected accuracy, then $f'(Pr, w_1) = f'(Pr, w_4) = Pr'''$, where $Pr'''(\{w_1\}) = Pr'''(\{w_2\}) = \frac{1}{6}$, $Pr'''(\{w_3\}) = 0$ and $Pr'''(\{w_4\}) = \frac{2}{3}$; and $f'(Pr, w_2) = f'(Pr, w_3) = Pr''''$, where $Pr''''(\{w_1\}) = Pr''''(\{w_2\}) = \frac{1}{6}$, $Pr''''(\{w_3\}) = \frac{2}{3}$

and $Pr''''(\{w_4\}) = 0$.

Bronfman's additional constraint appears fairly innocuous. As a general constraint, it might be questionable, but it's hard to see why there could not be at least some cases of the type he describes. Moreover, his example shows that even very severe constraints on the set of available update functions will not be sufficient to vindicate conditioning in the planning framework. About the strictest constraints on availability I can imagine are:

> **No unknown certainties**: If $Pr(p) < 1$ and $E(w) \not\subset p$, then $f(Pr, w)(p) < 1$.
>
> **All certainties known**: If $E(w) \subseteq p$, then $f(Pr, w)(p) = 1$.

Jointly: $f(Pr, w)(p) = 1$ iff $E(w) \subseteq p$. These forbid 'evidential overreach' in the sense that we saw in the variant update rules: your evidence doesn't in general entail facts about what your total evidence is, or about your update dispositions, so will rule out any certainties in facts about that in the cases at hand. You can only be certain of things that your evidence guarantees. Yet what Bronfman's example shows is that *even this* will not stack the deck in favour of conditioning. For in that example, the alternative, expected accuracy-maximising posteriors satisfy both of these constraints, in addition to the Greaves and Wallace availability constraint. I find it hard to imagine further constraints that would not simply stipulate conditioning as the most expectedly-accurate rule — leading to the dialectical weakness I described above.

Bronfman's example makes me worry that the problem with accuracy arguments for conditioning in the planning framework cannot just be that they assume you will have information you in fact lack.

There are always other strategies for someone who wishes to tell an accuracy-conducive story about conditioning. A 'scorched-earth' move would be to reject

the planning framework altogether, and instead endorse an accuracy argument in the key of Leitgeb-Pettigrew-Gallow. Personally, I am not especially convinced of this approach. Moreover, it would be nice, in this scenario, to have some positive reasons for rejecting the planning framework. But I actually find the planning framework fairly attractive — certainly more attractive than the Leitgeb-Pettigrew-Gallow framework. I am even a little tempted by some of the stronger claims floated in connection with the planning framework — for instance, that absent reasons to change your mind, you should stick by your intentions — that might allow one to get a genuinely diachronic requirement to condition out of the planning framework.[24] In short, I like evidence externalism, I like strict propriety, and I wish I knew what to make of the foregoing.

5.7 Conclusion

Let me summarise. We've seen that given evidence externalism, the expectedly accurate update function will not be conditionalisation, and will be some update function that sacrifices certain evidential virtues. At first glance this is a bad problem for the hegemonic ambitions of the accuracy programme. With that said, I've argued that the problem may arise out of a more general tension between a condition known as strict propriety, that is not proprietary to the accuracy programme, and evidence externalism. As a result, the tension here may not afflict accuracy epistemology in particular. Instead, that tension might seem to suggest one of two things: an argument against evidence externalism, and in favour of evidence internalism, or, an argument against strict propriety, and by extension, the and accuracy programme. Either outcome is significant.

[24]Cf. Paul (2014); Pettigrew (2016a, pp. 206–208).

6 | Conclusion

Let me summarise the primary findings of this thesis.

There is a good case that the kind of accuracy coherence needed for outright belief generates much stronger requirements than has been thought: namely, that the set of your beliefs should be deductively cogent. The epistemology of higher-order belief that emerges from this kind of coherence is incomplete, but hits at least some minimal desiderata, such as forbidding, on grounds of accuracy, believing blindspots, or that your beliefs are a(n egregiously) bad guide to the truth.

The standard accounts of normative uncertainty found in ethics, transferred to the case of uncertainty over scoring rules in accuracy epistemology, resolve the gap in the accuracy argument for probabilism identified by Aaron Bronfman.

There is a tension in the accuracy-theoretic account of evidence management, centred on how we evaluate the accuracy of updates. If we do so locally, we get an attractive account of updating, but we fail to vindicate the truism that gathering evidence is epistemically good for you. If we do so globally, we vindicate the truism that gathering evidence is epistemically good for you, but at the cost of an unattractive theory of updating. It's not clear how this tension should be resolved, though possibly the most promising route involves pushing the burden of the problem onto evidence externalists rather than accuracy epistemologists.

The thesis points to a number of directions for future research.

Chapters one and two open up a number of questions. First, is there a natural way to evaluate different theories of belief dynamics given the theories of accuracy-coherence we explored there? I gestured in section 1.7.4 at some ways in which the belief-sensitive $\cap$-coherence might fit better with the orthodox AGM theory than the belief-insensitive W-coherence. However my discussion was highly cursory. Second, some of my arguments in chapter two appealed to dynamical properties of the frames we looked at. Spelling out these arguments properly

requires a much better technical understanding of what these dynamics look like. As we saw in chapters four and five, there's a rich accuracy-theoretic discussion about updating degrees of belief. The accuracy-theoretic picture of outright belief is still lagging slightly behind, here. A separate question concerns different kinds of accuracy functions? Could there be reasons to drop constraints on the accuracy function like separability or extensionality, and (how) does the picture change if we do? A third question concerns the relationship between the two formal approaches to our investigations of the higher-order belief principles. To what extent do the two frameworks align? Is one more philosophically attractive than the other? And so on. Another question concerns how the issues we looked at play out in the setting of W-coherence. We established some results with regards to W-coherence and the introspection principles in the appendix to chapter two. There is much scope to generalise these discussions, though in some respects it is technically challenging: since there are inconsistent, or merely-single premise closed W-coherent belief sets, we cannot always use Kripke frames to represent them, as they impose multi-premise closure and (when serial) consistency by fiat. An interesting project would be to extend the framework to neighborhood logic frames, which permit failures of multi-premise closure, and see in what ways (if any) W-coherence yields a different epistemology of higher-order belief to $\cap B$-coherence. Taking matters in a different direction, knowledge-theoretic, or 'gnostic' (Littlejohn & Dutant, 2021) approaches to this belief-dominance framework — where the value function has a score for known propositions, and (possibly distinct) scores for merely believed true propositions, and believed false propositions — remain under-explored.[1] In particular, it would be interesting to (try to) extend the $\cap B$-coherence trick to what you know, and see what closure conditions fall out of this.

[1] Julien Dutant and Branden Fitelson also have some unpublished work on this.

Chapter three exploits the standard accounts of normative uncertainty in the ethics literature. Moreover, as I argued (admittedly, not rigorously), there's a reasonable case that the normative externalist and maximise expected choiceworthiness views exhaust the space of natural rules. With that said, I think there is room to expand the central results of the chapter. One way of thinking about the uncertainty literature is as involving questions about judgment aggregation, where the judgments involved are different (epistemic/moral) value functions.[2] In that case, it ought to be possible to ask much more general questions about the conditions under which these accuracy functions can be aggregated to yield an accuracy function that is continuous and strictly proper.

Chapters four and five raise questions about theories of updating on the accuracy framework that, as I mentioned at the end of the previous chapter, I still do not know what to make of. I would sorely like a clearer picture of what's going on here, though I have thought and talked about the problem enough to think that there is, at least, no completely straightforward answer. It was recently pointed out to me by an anonymous conference referee that most of the arguments in chapter five do not rely on any quantitative features of the update functions, and it may therefore be possible to generate stronger, purely qualitative versions of the problems I discussed.

As always, there is work to do.

[2]I owe this point to Matt Mandelkern.

References

Ahmed, A., & Salow, B. (2019). Don't Look Now. *British Journal for the Philosophy of Science*, *70*(2), 327–350.

Arpaly, N. (2002). *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford University Press.

Askell, A. (2018). Epistemic Consequentialism and Epistemic Enkrasia. *Epistemic Consequentialism*, 290–309.

Ayer, A. (1957). The Conception of Probability as a Logical Relation. In S. Körner (Ed.), *Observation and Interpretation in the Philosophy of Physics, with Special Reference to Quantum Mechanics Proceedings* (pp. 12–30). Dover Publications.

Baltag, A., & Renne, B. (2016). Dynamic Epistemic Logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/win2016/entries/dynamic-epistemic/`.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3.

Briggs, R. A., & Pettigrew, R. (2020). An Accuracy-Dominance Argument for Conditionalization. *Noûs*, *54*(1), 162–181.

Bronfman, A. (n.d.). *A Gap In Joyce's Argument for Probabilism.*

Bronfman, A. (2014). Conditionalization and Not Knowing That One Knows. *Erkenntnis*, *79*(4), 871–892.

Broome, J. (2004). *Weighing Lives*. Oxford University Press.

Broome, J. (2007). Wide or Narrow Scope? *Mind*, *116*(462), 359–370.

Brown, J. (2018). *Fallibilism: Evidence and Knowledge*. Oxford University

Press.

Buchak, L. (2014). Belief, Credence, and Norms. *Philosophical Studies*, *169*(2), 1–27.

Caie, M. (2013). Rational Probabilistic Incoherence. *Philosophical Review*, *122*(4), 527–575.

Campbell-Moore, C., & Levinstein, B. (2021). Strict Propriety Is Weak. *Analysis*, *81*(1), 8–13.

Campbell-Moore, C., & Salow, B. (2020a). Accurate Updating for the Risk Sensitive. *British Journal for the Philosophy of Science*, axaa006.

Campbell-Moore, C., & Salow, B. (2020b). Avoiding risk and avoiding evidence. *Australasian Journal of Philosophy*, *98*(3), 495–515. doi: 10.1080/00048402.2019.1697305

Carr, J. R. (2015). Epistemic Expansions. *Res Philosophica*, *92*(2), 217–236.

Carr, J. R. (2017). Epistemic Utility Theory and the Aim of Belief. *Philosophy and Phenomenological Research*, *95*(3), 511–534.

Carter, S. (2019). Higher Order Ignorance Inside the Margins. *Philosophical Studies*, *176*(7), 1789–1806.

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

Clarke, R. (2013). Belief Is Credence One (in Context). *Philosophers' Imprint*, *13*, 1–18.

Clifford, W. K. (1877). The ethics of belief. In *The Ethics of Belief and Other Essays* (pp. 70–97). Amherst, New York: Prometheus Books.

Cohen, M. (2021). Opaque Updates. *Journal of Philosophical Logic*, *50*(3), 447–470.

Das, N. (2019). Accuracy and ur-prior conditionalization. *Review of Symbolic Logic*, *12*(1), 62–96. doi: 10.1017/s1755020318000035

Das, N. (2022). Externalism and Exploitability. *Philosophy and Phenomenological Research*, 101–128.

Das, N. (in press). The Value of Biased Information. *British Journal for the Philosophy of Science*, 1–40. doi: 10.1093/bjps/axaa003

de Finetti, B. (1937). La Prévision: Ses Lois Logiques, Ses Sources Subjectives. *Annales de l'Institut Henri Poincaré*, *17*, 1–68.

DeRose, K. (1992). Contextualism and Knowledge Attributions. *Philosophy and Phenomenological Research*, *52*(4), 913–929.

Dorst, K. (2019). Lockeans Maximize Expected Accuracy. *Mind*, *128*(509), 175–211.

Dorst, K. (2020). Evidence: A Guide for the Uncertain. *Philosophy and Phenomenological Research*, *100*(3), 586–632.

Dorst, K., Levinstein, B. A., Salow, B., Husic, B. E., & Fitelson, B. (2021). Deference Done Better. *Wiley: Philosophical Perspectives*, *35*(1), 99–150.

Dorst, K., & Mandelkern, M. (in press). Good Guesses. *Philosophy and Phenomenological Research*. doi: 10.1111/phpr.12831

Easwaran, K. (2016). Dr. Truthlove Or: How I Learned to Stop Worrying and Love Bayesian Probabilities. *Noûs*, *50*(4), 816–853.

Elga, A. (2007). How to Disagree About How to Disagree. In T. Warfield & R. Feldman (Eds.), *Disagreement* (pp. 175–186). Oxford, UK: Oxford University Press.

Feldman, R. (1995). In Defence of Closure. *Philosophical Quarterly*, *45*(181), 487–494.

Fitelson, B., & Easwaran, K. (2015). Accuracy, coherence, and evidence. *Oxford Studies in Epistemology*, *5*, 61–96.

Foley, R. (2009). Beliefs, Degrees of Belief, and the Lockean Thesis. In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of Belief* (pp. 37–47). Springer.

Friedman, J. (2013). Suspended Judgment. *Philosophical Studies*, *162*(2), 165–181.

Gallow, J. D. (2019). Learning and Value Change. *Philosophers' Imprint*, *19*, 1–22.

Gallow, J. D. (2021). Updating for Externalists. *Noûs*, *55*, 487–516.

Gibbard, A. (2008). Rational Credence and the Value of Truth. *Oxford Studies in Epistemology*, *2*, 143–164.

Goldman, A. (1999). *Knowledge in a Social World*. Oxford, England: Oxford University Press.

Goldman, A. (2009). Williamson on Knowledge and Evidence. In P. Greenough, D. Pritchard, & T. Williamson (Eds.), *Williamson on knowledge* (pp. 73–91). Oxford University Press.

Good, I. J. (1967). On the Principle of Total Evidence. *British Journal for the Philosophy of Science*, *17*(4), 319–321. doi: 10.1093/bjps/17.4.319

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

Graves, P. R. (1989). The Total Evidence Theorem for Probability Kinematics. *Philosophy of Science*, *56*(2), 317–324.

Greaves, H. (2013). Epistemic Decision Theory. *Mind*, *122*(488), 915–952.

Greaves, H. (2017). Population Axiology. *Philosophy Compass*, *12*(11), 1–15.

Greaves, H., & Wallace, D. (2005). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, *115*(459), 607–632.

Greco, D. (2014). Could KK Be OK? *Journal of Philosophy*, *111*(4), 169–197.

Gustafsson, J. E., & Torpman, O. (2014). In Defence of My Favourite Theory. *Pacific Philosophical Quarterly*, *95*(2), 159–174.

Harman, E. (2015). The Irrelevance of Moral Uncertainty. *Oxford Studies in Metaethics*, *10*, 53–79.

Hawthorne, J. (2003). *Knowledge and Lotteries*. Oxford University Press.

Hawthorne, J. (2005). The Case for Closure. In M. Steup & E. Sosa (Eds.), *Contemporary Debates in Epistemology* (pp. 26–43). Blackwell.

Hawthorne, J., Rothschild, D., & Spectre, L. (2016). Belief is Weak. *Philosophical Studies*, *173*(5), 1393–1404.

Hedden, B. (2016). Does MITE Make Right? Decision-Making Under Normative Uncertainty. *Oxford Studies in Metaethics*, *11*, 102–128.

Hempel, C. (1962). Deductive-Nomological vs. Statistical Explanation. *Minnesota Studies in the Philosophy of Science*, *3*, 98–169.

Hewson, M. (2020). Accuracy Monism and Doxastic Dominance: Reply to Steinberger. *Analysis*, *80*, 450–456.

Hewson, M. (2021). Accurate Believers Are Deductively Cogent. *Noûs*.

Horowitz, S. (2017). Epistemic Value and the Jamesian Goals. In J. D. K. Ahlstrom-Vij (Ed.), *Epistemic consequentialism* (p. 269-289). Oxford University Press.

Huemer, M. (2001). *Skepticism and the veil of perception*. Lanham: Rowman & Littlefield.

Hughes, N. (2021). Epistemology Without Guidance. *Philosophical Studies*, *179*(1), 163–196.

Hutchison, K. (1999). What Are Conditional Probabilities Conditional Upon? *British Journal for the Philosophy of Science*, *50*(4), 665–695.

Jackson, F. (1991). Decision-Theoretic Consequentialism and the Nearest and Dearest Objection. *Ethics*, *101*(3), 461–482.

James, W. (1897). *The will to believe: And other essays in popular philosophy*. New York: Longmans, Green and Co.

Jeffrey, R. C. (1965). *The logic of decision*. New York, NY, USA: University of Chicago Press.

Joyce, J. (1998). A Nonpragmatic Vindication of Probabilism. *Philos. Sci*, *65*.

Joyce, J. (2009). Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of belief* (pp. 263–297). Synthese.

Joyce, J. (2018). Remarks on Richard Pettigrew's Accuracy and the Laws of Credence. *Philosophy and Phenomenological Research*, *96*(3), 755–762.

Kaplan, M. (1996). *Decision Theory as Philosophy*. Cambridge University Press.

Kelly, T. (2013). Evidence Can Be Permissive. In M. Steup & J. Turri (Eds.), *Contemporary debates in epistemology* (p. 298). Blackwell.

Krauss, S. F. (2017). Lying, Risk and Accuracy. *Analysis*, *77*(4), 726–734. doi: 10.1093/analys/anx105

Kyburg, H. E. (1961). *Probability and the Logic of Rational Belief*. Middletown, Conn., Wesleyan University Press.

Lasonen-Aarnio, M. (2015). New Rational Reflection and Internalism About Rationality. *Oxford Studies in Epistemology*, *5*.

Leitgeb, H. (2017). *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford University Press.

Leitgeb, H., & Pettigrew, R. (2010a). An Objective Justification of Bayesianism I: Measuring Inaccuracy. *Philosophy of Science*, *77*(2), 201–235.

Leitgeb, H., & Pettigrew, R. (2010b). An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy. *Philosophy of Science*, *77*(2), 236–272.

Levi, I. (1961). Decision Theory and Confirmation. *Journal of Philosophy*, *58*(21), 614–625. doi: 10.2307/2023537

Levi, I. (1965). Deductive Cogency in Inductive Inference. *Journal of Philosophy*, *62*(3), 68–77.

Levi, I. (1967). *Gambling with Truth: An Essay on Induction and the Aims of*

Science. MIT Press.

Levinstein, B. A. (2012). Leitgeb and Pettigrew on Accuracy and Updating. *Philosophy of Science*, *79*(3), 413–424.

Levinstein, B. A. (2017a). Accuracy Uncomposed: Against Calibrationism. *Epistemе*, *14*(1), 59–69.

Levinstein, B. A. (2017b). Permissive Rationality and Sensitivity. *Philosophy and Phenomenological Research*, *94*(2), 342–370.

Levinstein, B. A. (in press). Accuracy, Deference, Chance. *Philosophical Review*.

Lewis, D. (1971). Immodest Inductive Methods. *Philosophy of Science*, *38*(1), 54–63.

Lewis, P. J., & Fallis, D. (2019). Accuracy, Conditionalization, and Probabilism. *Synthese*, *198*, 4017–4033.

Littlejohn, C. (2014). *Justification and the truth-connection*. Cambridge University Press.

Littlejohn, C. (in press). Moore's Paradox and Assertion. In *Oxford handbook of assertion.* Oxford University Press.

Littlejohn, C., & Dutant, J. (2021). Defeaters as Indicators of Ignorance. In M. Simion & J. Brown (Eds.), *Reasons, Justification, and Defeat* (pp. 223–246). Oxford University Press.

Lockhart, T. (2000). *Moral Uncertainty and its Consequences*. Oxford University Press.

Loets, A. J. (2022). Choice Points for a Theory of Normality. *Mind*, *131*, 159–191.

MacAskill, W. (2014). *Normative uncertainty* (Unpublished doctoral dissertation). University of Oxford.

MacAskill, W. (2016). Smokers, Psychos, and Decision-Theoretic Uncertainty. *Journal of Philosophy*, *113*, 425–445.

MacAskill, W., Bykvist, K., & Ord, T. (2020). *Moral Uncertainty*. Oxford University Press.

MacAskill, W., & Ord, T. (2020). Why Maximize Expected ChoiceWorthiness? *Noûs*, *54*(2), 327–353.

Magidor, O. (2018). How Both You and the Brain in a Vat Can Know Whether or Not You Are Envatted. *Aristotelian Society Supplementary Volume*, *92*(1), 151–181.

Maher, P. (2002). Joyce's Argument for Probabilism. *Philosophy of Science*, *69*(1), 73–81.

Makinson, D. C. (1965). The Paradox of the Preface. *Analysis*, *25*(6), 205.

Mitova, V. (2015). Truthy Psychologism About Evidence. *Philosophical Studies*, *172*(4), 1105–1126.

Moore, G. E. (1942). A Reply to My Critics. In P. A. Schilpp (Ed.), *The Philosophy of G. E. Moore.* Open Court.

Moss, S. (2018). *Probabilistic Knowledge*. Oxford University Press.

Oddie, G. (1997). Conditionalization, Cogency, and Cognitive Value. *British Journal for the Philosophy of Science*, *48*(4), 533–541.

Parfit, D. (1988). What We Together Do. *Unpublished manuscript*.

Paul, S. K. (2014). Diachronic Incontinence is a Problem in Moral Philosophy. *Inquiry: An Interdisciplinary Journal of Philosophy*, *57*(3), 337–355.

Pettigrew, R. (2011). An Improper Introduction to Epistemic Utility Theory. In H. de Regt, S. Okasha, & S. Hartmann (Eds.), *Proceedings of epsa: Amsterdam '09* (pp. 287–301). Springer.

Pettigrew, R. (2016a). *Accuracy and the Laws of Credence*. Oxford University Press UK.

Pettigrew, R. (2016b). Jamesian Epistemology Formalised: An Explication of 'the Will to Believe'. *Episteme*, *13*(3), 253–268.

Pettigrew, R. (2017). Epistemic Utility and the Normativity of Logic. *Logos and Episteme*, *8*(4), 455–492.

Pettigrew, R. (2018). The Population Ethics of Belief: In Search of an Epistemic Theory X. *Noûs*, *52*(2), 336–372.

Pettigrew, R. (2019). Epistemic Utility Arguments for Probabilism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2019 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/win2019/entries/epistemic-utility/`.

Pettigrew, R. (2020). *The Dutch Book Arguments*. Cambridge, UK: Cambridge University Press.

Pettigrew, R. (2022). Accuracy-First Epistemology Without Additivity. *Philosophy of Science*, *89*, 128–151.

Pettigrew, R. (in press-a). Bayesian Updating When What You Learn Might Be False. *Erkenntnis*, 1–16. doi: 10.1007/s10670-020-00356-8

Pettigrew, R. (in press-b). *Epistemic Risk and the Demands of Rationality*. Oxford University Press.

Portmore, D. W. (2007). Consequentializing Moral Theories. *Pacific Philosophical Quarterly*, *88*(1), 39–73.

Predd, J. B., Seiringer, R., Lieb, E. H., Osherson, D. N., Poor, H. V., & Kulkarni, S. R. (2009). Probabilistic coherence and proper scoring rules. *IEEE Transactions on Information Theory*, *55*(10), 4786–4792.

Pryor, J. (2000). The skeptic and the dogmatist. *Noûs*, *34*(4), 517–549. doi: 10.1111/0029-4624.00277

Pryor, J. (2004). What's Wrong with Moore's Argument? *Philosophical Issues*, *14*(1), 349–378.

Ramsey, F. P. (1926). Truth and Probability. In A. Eagle (Ed.), *Philosophy of*

probability: Contemporary readings (pp. 52–94). Routledge.

Ramsey, F. P. (1990). Weight or the Value of Knowledge. *British Journal for the Philosophy of Science*, *41*(1), 1–4.

Rothschild, D. (2020). What it takes to believe. *Philosophical Studies*, *177*(5), 1345–1362. doi: 10.1007/s11098-019-01256-6

Salow, B. (2018). The Externalist's Guide to Fishing for Compliments. *Mind*, *127*(507), 691–728.

Salow, B. (in press). The Value of Evidence. In M. Lasonen-Aarnio & C. Littlejohn (Eds.), *The routledge handbook for the philosophy of evidence* (pp. 1–18).

Savage, L. J. (1954). *The foundations of statistics*. Wiley Publications in Statistics.

Schellenberg, S. (2013). Experience and evidence. *Mind*, *122*(487), 699–747. doi: 10.1093/mind/fzt088

Schoenfield, M. (2014). Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief. *Noûs*, *48*(2), 193–218.

Schoenfield, M. (2016). An Accuracy Based Approach to Higher Order Evidence. *Philosophy and Phenomenological Research*, 690–715.

Schoenfield, M. (2017). Conditionalization Does Not Maximize Expected Accuracy. *Mind*, *126*(504), 1155–1187.

Schultheis, G. (2018). Living on the Edge: Against Epistemic Permissivism. *Mind*, *127*(507), 863–879.

Schwitzgebel, E. (2006). The Unreliability of Naive Introspection. *Philosophical Review*, *117*(2), 245–273.

Sepielli, A. (2009). What to Do When You Don't Know What to Do. *Oxford Studies in Metaethics*, *4*, 5–28.

Sepielli, A. (2016). Moral uncertainty and fetishistic motivation. *Philosophical Studies*, *173*(11), 2951–2968.

Shear, T., & Fitelson, B. (2019). Two Approaches to Belief Revision. *Erkenntnis*, *84*(3), 487–518.

Skipper, M. (2021). Belief Gambles in Epistemic Decision Theory. *Philosophical Studies*, *178*(2), 407–426.

Skyrms, B. (1980). *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*. Yale University Press.

Smithies, D. (2019). *The epistemic role of consciousness*. New York, USA: Oxford University Press.

Sorensen, R. A. (1988). *Blindspots*. Oxford, England: Oxford University Press.

Stanley, J. (2005). *Knowledge and Practical Interests*. Oxford University Press.

Steel, R. (2018). Anticipating Failure and Avoiding It. *Philosophers' Imprint*, *18*.

Steinberger, F. (2019). Accuracy and Epistemic Conservatism. *Analysis*, *79*, 658–669.

Sturgeon, S. (2008). Reason and the Grain of Belief. *Noûs*, *42*(1), 139–165.

Sturgeon, S. (2019). Epistemology, Pettigrew Style: A Critical Notice of Accuracy and the Laws of Credence, by Richard Pettigrew. *Mind*, *128*(512), 1319–1336.

Topey, B. (in press). Best Laid Plans: Idealization and the Rationality-Rccuracy Bridge. *British Journal for the Philosophy of Science*. doi: https://doi.org/10.1086/718275

van Fraassen, B. C. (1984). Belief and the Will. *Journal of Philosophy*, *81*(5), 235–256.

Weatherson, B. (2019). *Normative Externalism*. Oxford, UK: Oxford University Press.

Wedgwood, R. (2002). The Aim Of Belief. *Noûs*, *36*, 267–297.

White, R. (2005). Epistemic Permissiveness. *Philosophical Perspectives*, *19*(1), 445–459.

White, R. (2006). Problems for Dogmatism. *Philosophical Studies*, *131*(3), 525–557.

Williams, J. R. G. (2015). Accuracy, Logic, and Degree of Belief. In C. Caret & O. Hjortland (Eds.), *Foundations of Logical Consequence.* Oxford University Press.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.

Williamson, T. (2014). Very Improbable Knowing. *Erkenntnis*, *79*(5), 971–999.