Logical Disagreement

Logical Disagreement

An Epistemological Study

by Frederik J. Andersen

Thesis for the degree of Doctor of Philosophy
Thesis advisors: Greg Restall, Franz Berto, Jessica Brown

Funding information: the Niels Bohr Foundation (Royal Danish Academy of Sciences) and the Royal Institute of Philosophy (London)

University of St Andrews, Department of Philosophy

Contents

Nothing you write is ever as bad as you fear or as good as you hope.

Abstract

While the epistemic significance of disagreement has been a popular topic in epistemology for at least a decade, little attention has been paid to *logical* disagreement. This monograph is meant as a remedy. The text starts with an extensive literature review of the epistemology of (peer) disagreement and sets the stage for an epistemological study of logical disagreement. The guiding thread for the rest of the work is then three distinct readings of the ambiguous term 'logical disagreement'. Chapters 1 and 2 focus on the *Ad Hoc Reading* according to which logical disagreements occur when two subjects take incompatible doxastic attitudes toward a specific proposition in or about logic. Chapter 2 presents a new counterexample to the widely discussed Uniqueness Thesis. Chapters 3 and 4 focus on the *Theory Choice Reading* of 'logical disagreement'. According to this interpretation, logical disagreements occur at the level of entire logical theories rather than individual entailment-claims. Chapter 4 concerns a key question from the philosophy of logic, viz., how we have epistemic justification for claims about logical consequence. In Chapters 5 and 6 we turn to the *Akrasia Reading*. On this reading, logical disagreements occur when there is a mismatch between the deductive strength of one's background logic and the logical theory one prefers (officially). Chapter 6 introduces *logical akrasia* by analogy to epistemic akrasia and presents a novel dilemma. Chapter 7 revisits the epistemology of peer disagreement and argues that the epistemic significance of central principles from the literature are at best deflated in the context of logical disagreement. The chapter also develops a simple formal model of *deep disagreement* in Default Logic, relating this to our general discussion of logical disagreement. The monograph ends in an epilogue with some reflections on the potential epistemic significance of *convergence* in logical theorizing.

Acknowledgements

I'm grateful to many others for support in writing this thesis over the course of four years. First and foremost, Greg Restall, who has been my primary supervisor for the last two of them. Greg has without any doubt been the greatest inspiration in my academic career thus far. I see Greg as nothing short of an intellectual force of nature, but at the same time he's a very kind and gentle person. I'm entirely grateful to him for being courageous enough to take over the main supervision of my project so late in the process (and during a pandemic!). Additionally I'm extremely thankful to Jessica Brown. Jessica was my main supervisor for two years under very difficult and unprecedented circumstances, but nonetheless she kept razor sharp, supportive, and available. Jessica has this special ability to make all your (bad) arguments fall apart while smiling at you. For this and much else I admire her deeply. I'm also grateful for all the help and support I've received from the last member of my supervisor team, Franz Berto. I'm merely preaching to the choir by stating that Franz is the coolest philosopher around these days. It's always inspiring to see him in action because you can really tell that he loves his job. So while I've always feared that I'm one of those academics he dislikes because they take themselves a bit too seriously, I'm proud to say that Franz has been a constant reminder of why I got interested in philosophy to begin with. The philosophy business is simply too tough a game if one forgets to have fun while playing.

Another special thanks goes to Gillian Russell, who provided invaluable help and super annoying comments for the fourth chapter of the dissertation.

Further, I need to thank the many other philosophers and logicians who have helped me in different ways during the PhD-period. The list includes, but is not limited to: Aaron Cotnoir, Adriano Angelucci, Alessandro Rossi, Andrea Oliani, Andreas Christiansen, Angie O'Sullivan, Ben Martin, Bjørn Gunnar Hallsson, Camille Fouche, Catarina Dutilh Novaes, Claire Field, Dario Mortini, Diego Tajer, Francisca Silva, Henrik Sherling, Hugh Reilly, Jace Snodgrass, James Skinner, Jann Paul Engler, Justin Snedegar, Katharina Bernhard, Katherine Hawley, Kent Hurtig, Klemens Kappel, Lara Jost, Leon Commandeur, Lilith Newton, Makan Nojoumian, Martin Lipman, Matteo Nizzardo, Michael Potter, Mikkel Gerken, Nick Kuespert, Nikolaj J. L. L. Pedersen, Nils Holtug, Ole Hjortland, Paolo Savino, Patrick Blackburn, Rasmus Overmark, Rasmus K. Rendsvig, Ross Patrizio, Roy Sorensen, Sandy Goldberg, Sophie Nagler, Thomas Radriamahazaka, Thor Grünbaum, Victor Lange, Vincent F. Hendricks, Viviane Fairbank, and Walter B. Pedriali.

Finally, I thank my friends and family outwith the academic community for their unwavering support. You know who you are.

Fred Andersen
St Andrews, September 2023

Introduction

While the epistemic significance of disagreement has been a popular topic in mainstream epistemology for over a decade now, surprisingly little attention has been paid to *logical* disagreement in particular. This monograph is meant as a remedy. By focusing on disagreements *in* and *about* logic, it will be able to fill an important gap in the contemporary literature.

The monograph consists of an introduction, three preamble chapters, and four research chapters followed by a short epilogue and two technical appendices.[1]

In the present introduction the reader will find an extensive literature review of the epistemology of (peer) disagreement aiming to set the stage for a thorough epistemological study of *logical disagreement.* The guiding thread for the rest of the work will then be three distinct readings of the ambiguous term 'logical disagreement'.

The focal point of Chapters 1 and 2 is the *Ad Hoc Reading* of 'logical disagreement'. According to this interpretation, logical disagreements occur when two (or more) subjects take incompatible doxastic attitudes toward a specific proposition in or about logic, e.g., whether Modus Ponens is a valid inference. The second chapter is based on the paper *Uniqueness and Logical Disagreement* (Andersen, 2020, 2023b), which discusses the *Uniqueness Thesis* (a core thesis in the epistemology of disagreement). After presenting the Uniqueness Thesis and clarifying relevant terms, a novel counterexample to the thesis is introduced. The counterexample involves logical disagreement under the Ad Hoc Reading. Several objections to the counterexample are then considered, and it's argued that the best responses to the counterexample all undermine the initial motivation for uniqueness in the relevant sense.

In Chapters 3 and 4 our focus will be on the *Theory Choice Reading* of 'logical disagreement' instead. This reading of 'logical disagreement' is one that happens to be in vogue at the time of writing. On this interpretation, genuine logical disagreements occur at the level of entire logical theories rather than individual entailment-claims. Chapter 4 is based on the paper *Countering Justification Holism in the Epistemology of Logic: The Argument from Pre-Theoretic Universality* (Andersen, 2023a) and it concerns a key question from the philosophy

[1] The purpose of the preambles will be explained to the reader in due course.

of logic, viz., how we have epistemic justification for claims about logical consequence (assuming that we have such justification at all). Justification holism asserts that claims of logical consequence can only be justified in the context of an entire logical theory, e.g., classical, intuitionistic, paraconsistent, paracomplete etc. So, according to holism, claims of logical entailment cannot be atomistically justified as isolated statements, independently of theory choice. At present there is a developing interest in—and endorsement of—justification holism due to the revival of an abductivist approach to the epistemology of logic. The fourth chapter gives an argument against holism by establishing a foundational entailment-sentence of deduction which is justified independently of theory choice and outside the context of a whole logical theory.

In Chapters 5 and 6 we'll turn to the *Akrasia Reading* of 'logical disagreement'. On this reading, logical disagreements occur when there is a mismatch between the deductive strength of one's background logic—i.e., the logic one uses to prove metatheoretic results such as soundness and completeness—and the logical theory one prefers (officially). The sixth chapter evolved from the paper *Logical Akrasia* (Andersen, 202X) and has two main aims. First, it introduces the novel concept *logical akrasia* by analogy to epistemic akrasia; second, it presents a dilemma based on logical akrasia. From a case involving the consistency of Peano Arithmetic and Gödel's Second Incompleteness Theorem it's shown that either we must be agnostic about the consistency of Peano Arithmetic or akratic in our logical theorizing. If successful, the initial sections of the chapter will draw attention to an interesting akratic phenomenon which has not received much attention in the literature on akrasia (although it has been discussed by logicians in different terms); while the final sections try to underscore the pertinence and persistence of akrasia in logic (by appeal to Gödel's seminal work). The chapter eventually concludes by suggesting a way of translating the dilemma of logical akrasia into a case of regular epistemic akrasia; and further how one might try to escape the dilemma when it's framed this way.

Finally, in Chapter 7, we take stock. Chapter 7 explores the epistemic significance of logical disagreement in light of the initial literature review and the three different interpretations of logical disagreement from previous chapters. In general outline the chapter consists of two parts. First and foremost, it revisits the epistemology of peer disagreement and argues that the epistemic significance of central principles from the literature—e.g., Epistemic Peerhood and Independence—are at best deflated when applied in the context of logical disagreement. The cumulative outcome is thus a skeptical pressure against sweeping answers to the Doxastic Disagreement Question: *What is the epistemically rational response to cases where one disagrees with an epistemic peer as to whether proposition $\langle p \rangle$ is true?* Since it's not even possible to give a normatively satisfying answer in the cases where $\langle p \rangle$ happens to be a logical proposition, we can't give a completely general answer either.[2] Secondly, the chapter develops a simple

[2] We'll use the angle bracket-notation throughout the monograph to indicate when we are interested in the *proposition* expressed by a given declarative sentence such as p rather than the sentence itself.

formal model of paradigmatic *deep disagreement* in a refined Horty-style Default Logic and compares the result with some obvious competitors. As will become clear, the simple Default Logic-model fares quite well in comparison to both Classical Propositional Logic and Subjective Bayesianism. Discussions of different formal models of deep disagreement are then related to our general discussion of logical disagreement, and it is concluded that *if* logical disagreements are structurally similar to deep disagreements, then how we ought to respond to logical disagreements—epistemically speaking—will depend on the legitimacy of *subjective* rankings of logical principles.

The monograph ends with some reflections on the interesting, and yet underexplored, epistemological flip side of logical disagreement, viz., the epistemic significance of *agreement* (or convergence) in logical theorizing. While the idea of convergence is familiar from the philosophy of science, there is almost no sustained discussion of the nature and epistemic significance of convergence in (the philosophy of) logic. Perhaps convergence in logic is epistemically significant because converging logical theories constitute independent methods of reasoning confirming the same results, e.g., that Modus Ponens is a deductively valid inference, and when independent methods confirm the same results, we have more reason to trust them. While initially appealing, there are serious challenges to this idea upon reflection. The first challenge lies in understanding what it means for methods of reasoning to be independent of each other on a generic level and explaining why convergence of such independent methods is epistemically significant. A second challenge is deciding how (or if at all) logical theories may be considered independent methods of reasoning, the convergence of which is epistemically significant. As will become clear, even if we can say something useful about the first of these challenges, the second is not a straightforward matter.

Note that the text also includes two technical appendices. The first concerns Michael G. Titelbaum's controversial Fixed Point Thesis and his No Way Out-argument in support of it. The appendix is included because Titelbaum's thesis would have very severe repercussions for the epistemology of disagreement if it were true. The second appendix gives the technical definitions of the default logic from *Reasons as Defaults* due to John F. Horty (2012). This appendix is included as we'll rely quite heavily on parts of Horty's formal framework when building our model of deep disagreement in Chapter 7.

1 The Epistemology of Disagreement

Since the year 2000 a host of epistemologists have shown great interest in the general question of how to respond to (peer) *disagreement* in an epistemically rational way.[3] Anthologies

[3] (Christensen, 2007, 2009, 2010, 2014, 2016; Cohen, 2013; Elga, 2005, 2007, 2010; Feldman, 2005a, 2006, 2009; Frances, 2008; Goldman, 2010; Grundmann, 2019; Kappel, 2012, 2019a; Kappel and Andersen, 2019; Kappel, 2021; Kelly, 2005, 2008, 2010, 2013; King, 2012; Lackey, 2010, 2013; Matheson, 2009, 2014, 2015; Pettit, 2006;

(Christensen and Lackey, 2013; Feldman and Warfield, 2010) and textbooks (Frances, 2014; Matheson, 2015) have been devoted to the topic, and (peer) disagreement is now considered among the central themes of social epistemology (Goldman and Whitcomb, 2011). This is all with good reason; disagreements are abundant in many intellectual areas of human life—e.g., natural science, religion, ethics, politics, art, aesthetics, mathematics, and logic—and frequently they constitute an important part of the deliberative processes that lead to our beliefs and actions.

The crux of the disagreement literature has been to answer what may be termed the *Doxastic Disagreement Question*: What is the epistemically rational response to cases, where one disagrees doxastically with an epistemic peer as to whether $\langle p \rangle$?[4] [5]

Generally construed, two agents (alternatively: subjects, reasoners, cognizers etc.)—call them 'S' and 'S^*'—disagree as to whether $\langle p \rangle$ when they take different doxastic attitudes toward $\langle p \rangle$. Assuming a *Tripartite View*, the possibility space of doxastic attitudes is exhausted by *belief that* $\langle p \rangle$; *disbelief that* $\langle p \rangle$; and *suspension of judgment with respect to* $\langle p \rangle$.[6] Most obviously, agents S and S^* are in disagreement with respect to $\langle p \rangle$ if S believes that $\langle p \rangle$ while S^* disbelieves that $\langle p \rangle$ (or vice versa). But it would also count as a case of disagreement on this view if one party (dis)believes $\langle p \rangle$ whereas the other suspends judgement.

Alternatively, one could take a *Subjective Probability View* asserting that S and S^* disagree as to whether $\langle p \rangle$ if they adopt different credences toward $\langle p \rangle$. For instance, if S holds credence 0.3 in $\langle p \rangle$ and S^* holds credence 0.8 toward the same proposition, then S and S^* are in a case of disagreement with respect to $\langle p \rangle$. Of course, there is a further question of how fine-grained such a view should be, e.g., would it also count as a disagreement if S held a credence of 0.812 in $\langle p \rangle$ while S^* held 0.813? And if so, would a highly fine-grained view be plausible considering the limited nature of human cognition? For the sake of this introduction, we'll not need to take a stance on this issue, but simply note that paying attention to the technical details of, for example, the granularity of subjective probabilities can be of utmost importance when modeling scenarios of disagreement.

Plantinga, 2000; Rosen, 2001; Ranalli, 2020, 2021; Sosa, 2010; Srinivasan and Hawthorne, 2013; Weatherson, 2007; Wedgwood, 2010, 2019; Wright, 2021)

[4] Here *epistemic* rationality should be contrasted with practical or instrumental rationality (cf. (Kolodny and Brunero, 2023)). *Doxastic* disagreement should be contrasted with action-disagreement (cf. (Frances and Matheson, 2019)). Note also that some would take the above formulation of the Doxastic Disagreement Question to be overly strong as they prefer to think of the question in terms of epistemic *permissibility* rather than *obligation*, see for example (Broncano-Berrocal and Simion, 2021).

[5] More on the central notion of *epistemic peerhood* below.

[6] Some have argued that the doxastic attitude of *disbelief that* $\langle p \rangle$ is non-equivalent to that of *believing the negation of* $\langle p \rangle$. See (Smart, 2021) for a recent argument. Unless otherwise stated we'll simply take *disbelief that* $\langle p \rangle$ and *believing the negation of* $\langle p \rangle$ as equivalent attitudes in what follows.

1.1 Standard Cases

To make things more vivid, let's look at some standard cases from the literature:

> **Restaurant.** Suppose that five of us go out to dinner. It's time to pay the check, so the question we're interested in is how much we each owe. We can all see the bill total clearly, we all agree to give a 20 percent tip, and we further agree to split the whole cost evenly, not worrying over who asked for imported water, or skipped desert, or drank more of the wine. I do the math in my head and become highly confident that our shares are \$43 each. Meanwhile, my friend does the math in her head and becomes highly confident that our shares are \$45 each. How should I react, upon learning of her belief? (Christensen, 2007, p. 193).
>
> **Horse Race.** You and I, two equally attentive and well-sighted individuals, stand side by side at the finish line of a horse race. The race is extremely close. At time t_0, just as the first horses cross the finish line, it looks to me as though Horse A has won the race in virtue of finishing slightly ahead of Horse B; on the other hand, it looks to you as though Horse B has won in virtue of finishing slightly ahead of Horse A. At time t_1, an instant later, we discover that we disagree about which horse has won the race. How, if at all, should we revise our original judgements on the basis of this new information? (Kelly, 2010, p. 113).

Now, what is the epistemic significance of the disagreements in **Restaurant** and **Horse Race** (if any)? Perhaps a reduction of one's initial confidence in the proposition under dispute is called for in both cases?

Conciliatory views hold that it's epistemically required for both parties of cases like **Restaurant** and **Horse Race** to reduce their initial confidence with respect to the target-proposition. *Steadfast* views, on the other hand, hold that there are cases of disagreement akin to **Restaurant** and **Horse Race**, where at least one side of the disagreement is rationally permitted—or perhaps even required—to maintain their initial confidence.

Notice that insofar we are solely interested in cases with two disagreeing subjects (which is standardly assumed), the only proper conciliatory response will be one where both sides reduce their initial confidence in the target-proposition. All other possible responses will count as steadfast.

1.2 Standard Idealizations

Already at this stage some readers might feel a bit skeptical about the entire epistemological setup. One potential worry is that standard cases like **Restaurant** and **Horse Race** are underdeveloped and too imprecise to make decisive assessments of.

Authors in the literature have adopted a number of idealizations about their central cases in order to accommodate such worries, i.e., they have tried to build models of disagreement that isolate *epistemically* relevant factor(s) and avoid confounders.

One common idealization has been to assume that the parties involved in disagreement are *epistemic peers*. Thomas Kelly, who introduced the term 'epistemic peerhood' (2005, p. 174) in the context of mainstream epistemology,[7] asserts that two agents are epistemic peers exactly when:

1. they are equals with respect to their familiarity with the evidence and arguments which bear on that question, and;
2. they are equals with respect to general epistemic virtues such as intelligence, thoughtfulness, and freedom from bias.[8]

This conjunctive definition poses the immediate question of how to interpret 'equals' in each of its conjuncts. Does being equals in (1) imply being identical? Or does it merely imply having sufficiently similar familiarity with the evidence and arguments which bear on the question in a weaker, non-identical sense? Does being equals in (2) imply processing the relevant evidence in identical ways? Or merely doing an equally good job when processing the evidence even if it's not done in exactly the same way?

For our current purposes there is no need to take a precise reading of Kelly's definition. It is, on the other hand, important to note that there are different, non-equivalent definitions of epistemic peerhood featuring in the literature. Adam Elga (2007, p. 499) holds that an agent is your epistemic peer:

> ...with respect to an about-to-be-judged claim if and only if you think that, conditional [on] the two of you disagreeing about the claim, the two of you are equally likely to be mistaken.

[7] Kelly borrowed the term from Gutting (1982).

[8] Kelly (2005) notes the familiar fact that, outside of a purely mathematical context, the standards of equality between two entities, along some dimension, are highly context-sensitive. Thus, whether two individuals count as epistemic peers will depend on the specific standards for epistemic peerhood within a given context. Similarly, whether two individuals count as 'the same height' will depend on the specific standards of measurement that are in play, see, e.g., (Lewis, 1979).

While this definition of epistemic peerhood is explicitly *conditional* on the disagreement at hand, one could alternatively:

> ...say that for you to regard another thinker as your 'epistemic peer'...is for you to attach an equally high unconditional probability to the hypothesis that that thinker will be right about that question as to the hypothesis that you will be right about that question. (Wedgwood, 2010, p. 236).

However construed, the peerhood-model of disagreement is meant to establish a certain kind of epistemic symmetry between the sides involved. The model aims to remove the confounding possibility of it being epistemic asymmetries between the disagreeing parties—rather than a supposed significance of the disagreement itself—driving our assessments of central cases.[9]

Another idealization which has been adopted frequently in the literature is *full disclosure*. Full disclosure is taken to be a certain state obtaining in disagreements when the involved parties:

> ...have thoroughly discussed the issues. They know each other's reasons and arguments, and that the other person has come to a competing conclusion after examining the same information. (Feldman, 2006, p. 419).

As was the case with peerhood, the aim of full disclosure is to free our models of disagreement from some potentially confounding factors. One confounder—supposedly removed by full disclosure obtaining—is that of discursive unclarity, i.e., the disagreement wouldn't simply disappear if one side were to repeat and clarify their arguments.

2 The Main Contenders

In this section we'll assume that the parties of the disagreement cases discussed are epistemic peers and that the condition of full disclosure is satisfied. As mentioned above, conciliatory views hold that it is epistemically required for both sides of peer disagreement to reduce their initial confidence in the proposition under dispute. In contrast, steadfast views claim that there are cases of peer disagreement where at least one side is epistemically permitted, or perhaps even required, to stay unaffected by the presence of disagreement.

One should notice the *universal* claim involved in defining conciliatory views, i.e., *all* cases of peer disagreement are cases where a reduced confidence in the target-proposition is called

[9] See Williamson (2017a) for some general perspectives on model-building in philosophy. We'll also motivate the use of (formal) models in philosophy in Chapter 7 when we construct a model of deep disagreement in a framework of default logic.

for on both sides of the dispute. Otherwise the view won't count as conciliatory. In contrast, steadfastness is simply the negation of conciliationism, i.e., an *existential* claim. It's sufficient for a view to count as steadfast that there exists a case of peer disagreement, where it's permissible for some side of the dispute to stick to their guns.

Below §2.1 will sketch the essentials of some prominent conciliatory views while §2.2 will do the same for some important steadfast views.

2.1 Conciliatory Views, Higher-Order Evidence, and the Independence Principle

Consider again **Restaurant** (cf. §1.1) and assume the idealizations of epistemic peerhood as well as full disclosure. Now, according to any genuine conciliatory view, the rational response to **Restaurant** is for my friend and me to each adopt a doxastic attitude closer to that of our interlocutor after learning about our dispute (Christensen, 2007). To be sure, the claim is that upon discovering our disagreement, it is no longer epistemically rational for any of us to hold our initial doxastic attitude with respect to the proposition under dispute (viz., the correct share of the restaurant bill). By assumption we are epistemic peers and thus equally likely to get things right, which seems to constitute a *defeater* of our initial attitudes.[10] As peers regarding the matter at hand we are equally good at calculating shares of bills in our heads; perhaps we even have fine track records to show for it. So, the fact that my friend has reached a result that differs from mine gives me some reason to think that her result is correct while my own is wrong.

This latter point—concerning the potential defeater which is generated by the fact that my peer and I disagree—can even be emphasized if we consider a version of **Restaurant** where ten epistemic peers disagree with me instead of just one. If we stipulate that each of my ten peers featuring in such an altered restaurant case has reached the same result independently of one another, then it would seem that I have an overwhelmingly strong reason to defer to the majority and become significantly less confident in my initial doxastic attitude.[11] But if disagreeing with ten epistemic peers has such a strong impact on what is epistemically rational

[10] To a first approximation *defeasibility* in epistemology refers to a doxastic attitude's liability to lose some positive epistemic status (or having this status downgraded in some way). More generally, defeasibility refers to a kind of epistemic vulnerability, the potential loss, reduction, or prevention, of some positive epistemic status (Sudduth, 2008). A *defeater* is, roughly speaking, a condition that actualizes this potential. A very common distinction in the literature on defeat—first drawn by Pollock in (1986)—is between *rebutting* and *undercutting* defeaters. According to Pollock, a rebutting defeater for some belief that $\langle p \rangle$ is a reason (in a broad sense) for believing the negation of $\langle p \rangle$, or for holding some proposition, $\langle q \rangle$, which is incompatible with $\langle p \rangle$ (Pollock, 1986, p. 38). An undercutting defeater for some belief that $\langle p \rangle$ is a reason (broadly construed) for no longer believing $\langle p \rangle$; not for believing its negation (Pollock, 1986, p. 39). For canonical work on defeaters in epistemology the reader should consult (Pollock, 1970, 1974, 1984, 1986, 1994). See also (Kelp, 2023) for a recent overview.

[11] Cf. the Condorcet Jury Theorem, see for instance (Dietrich and Spiekermann, 2022).

for me, why shouldn't disagreeing with just one peer have some impact in this regard?[12]

While all conciliatory views agree that the presence of peer disagreement constitutes a reason to reduce one's initial confidence in the target-proposition under dispute, they can diverge quite radically in their answers to the *Question of Significance*, i.e., the appropriate level of doxastic revision required in light of disagreeing with a peer. When considering this question, it can be helpful to draw a distinction between *first-order* and *higher-order evidence* (Christensen, 2010; Kelly, 2010; Matheson, 2009; Skipper and Steglich-Petersen, 2019; Horowitz, 2022). In **Restaurant** we can take the target-proposition under dispute to be the proposition expressed by the sentence *my share of the total bill is* $43. Let's label this sentence 'p'. Now, the first-order evidence for proposition $\langle p \rangle$ is the evidence that directly bears on the truth of it, e.g., the fact that the total bill is X (where X is simply a placeholder for some number of dollars). On the other hand, a proposition like the one expressed by the sentence *the fact that the total bill is X supports* $\langle p \rangle$ is a higher-order proposition regarding $\langle p \rangle$. Higher-order propositions are:

> ...propositions that concern an epistemic agent, her doxastic states or doxastic attitudes. They may concern the relation between evidence and first-order doxastic attitudes. For example, a higher-order proposition is that the medical resident's belief is well-supported by her evidence, or that I have processed certain evidence in a rational way in forming my doxastic attitude towards a first-order proposition. Thus, first-order and higher-order propositions are distinguished by their subject-matter. (Henderson, 2022, p. 515)

So, the fact that I have inferred that $\langle p \rangle$ follows from my first-order evidence (and I happen to have a good track record in the relevant domain) is higher-order evidence in cases like **Restaurant**.[13]

Here's another case to guide our intuitions about higher-order evidence:

> **The Pill**. Suppose I consider a mathematical problem on the basis of some evidence E. Suppose that E entails that p is the correct answer to the mathematical problem. After careful scrutiny I come to believe that the correct answer is p. I am then told by a credible source that without noticing I have ingested a reason-distorting pill that makes me completely unreliable with respect to those

[12] See (Kelly, 2010; Lackey, 2010, 2013) for discussions of similar cases. See also the discussion of Titelbaum's so-called "Crowdsourcing Argument" in Appendix 1 of the present monograph.

[13] For further discussion of what distinguishes higher-order evidence from other types of evidence, the reader should consult (Ye, 2023). In general outline Ye observes that the label 'higher-order evidence' has been used equivocally to refer to (a) evidence about the rationality of one's belief; (b) evidence about one's reliability; (c) evidence about what evidence one has; and (d) evidence about what one's evidence supports (Ye, 2023, p. 3). We'll return to Ye's recent work on higher-order evidence in Chapter 7.

> kinds of mathematical problems, though this is not in any way perceptible to me. (Kappel and Andersen, 2019, p. 1105)[14]

In this scenario we have that the credible testimony concerning my distorted reasoning abilities—caused by my ingestion of a certain drug—is higher-order evidence with respect to proposition $\langle p \rangle$. In other words, the testimony constitutes evidence that doesn't bear directly on the truth of $\langle p \rangle$, i.e., the correct answer to the math problem; instead it bears on the truth of the higher-order proposition expressed by the sentence *I'm reliable in assessing the available first-order evidence bearing on* $\langle p \rangle$.[15] [16]

Returning to the theme of conciliatory views answering the Question of Significance, one controversial answer is given by the so-called "*Equal Weight View*" (Elga, 2007; Matheson, 2009; Kelly, 2010). According to this view, I should assign the same epistemic weight to the higher-order evidence about my epistemic peer as I should to the higher-order evidence about myself. Assuming epistemic peerhood, I have no reason to favor the fact that I have inferred $\langle p \rangle$ from evidence E over the fact that my peer has inferred that $\langle p \rangle$ doesn't follow from E. After all, one of us has made an error and there is no special reason for me to suspect that my peer has erred rather than me. Thus, at least on one plausible interpretation, the Equal Weight View vindicates *splitting the difference* in cases of peer disagreement. As we should assign equal weight to the possibilities of each of us being right, each of us ought to move towards the middle point between our initial doxastic attitudes. For example, if subject S initially believed that $\langle p \rangle$ while subject S^* disbelieved $\langle p \rangle$, then both agents ought to suspend judgment as to whether $\langle p \rangle$ (assuming the standard version of a Tripartite View of doxastic attitudes, cf. §1). Similarly, if S's initial credence in $\langle p \rangle$ was 0.8 and S^*'s initial credence 0.2, then they ought to converge on 0.5 as their revised credence in $\langle p \rangle$ (assuming the standard version of a Subjective Probability View, cf. §1).

We won't assess the finer details nor the plausibility of the Equal Weight View here, but we'll consider various objections to conciliatory views of peer disagreement below.

Independence

Consider the following steadfast reaction to, say, **Horse Race**:

[14] Adapted from (Christensen, 2011, pp. 5-6).

[15] For more elaborate discussions of *higher-order defeat*, see for instance (Skipper, 2019; Ye, 2020). For a related discussion of *epistemic akrasia*, see for example (Field, 2019; Lasonen-Aarnio, 2014, 2020; Sliwa and Horowitz, 2015). Epistemic akrasia will also be a topic of discussion in Chapters 5 and 6 below.

[16] It's worth noting that my higher-order evidence in **The Pill** could itself be rebutted or undercut by further information (Pollock, 1970, 1986). Suppose, for example, I learn from a credible source that I have ingested a reasoning-distorting pill without noticing. Yet I'm also told, by another credible source, that I happen to be one of very few people on whom the active ingredient has no effect. In this case I have received evidence undercutting my higher-order evidence.

> In response to our disagreement as to whether $\langle p \rangle$ (based on our common evidence E), you said not-$\langle p \rangle$ while I said $\langle p \rangle$, so you must be mistaken about the issue at hand and you can't count as my peer after all.[17]

As has been noted in the literature, there is something dubious and question-begging about this steadfast line of reasoning (Christensen, 2009; Elga, 2007; Feldman, 2006). Christensen has proposed a conciliatory antidote in the following form:

> *The Independence Principle*. In evaluating the epistemic credentials of another's expressed belief about p, in order to determine how (or whether) to modify my own belief about p, I should do so in a way that doesn't rely on the reasoning behind my initial belief that p (Christensen, 2009).[18]

What this principle seems to imply is that in cases like **Horse Race** one cannot rely on the fact of having concluded $\langle p \rangle$ on the basis of evidence E in one's assessment of the epistemic significance of disagreement. According to Christensen, one ought to "bracket" one's initial reasoning behind believing that $\langle p \rangle$, when trying to evaluate the epistemic position of oneself versus the ditto of one's interlocutor.[19] So, in a certain sense, a legitimate epistemic reason to downgrade my assessment of your epistemic credentials needs to be *independent* of the dispute at hand. In the version of **Horse Race** stated above, I have no independent reason of that kind and thus I should consider you my epistemic peer, and I should revise my confidence in $\langle p \rangle$ accordingly. Of course, one could easily alter the case in such a way that you were severely drugged, sleep deprived, or otherwise epistemically impaired; and based on such alternations of **Horse Race** it is plausible to suggest that I would have dispute-independent reasons to think that my own epistemic position is superior to yours (because the intoxication, sleep deprivation etc. could have had a negative influence on the reliability of your vision and judgement); but as the case is stated here it seems that one should indeed conciliate.

Now, while the Independence Principle is at least *prima facie* plausible and motivated by off-putting intuitions concerning question-begging, the principle remains controversial. For a more detailed discussion, the reader can consult, e.g., (Christensen, 2009, 2011; Lord, 2014; Moon, 2018; Sosa, 2010). Further, we'll consider the plausibility of the Independence Principle again in the context of logical disagreement in Chapter 7.

[17] Notice how this response hinges on whether we accept a conditional definition of epistemic peerhood, cf. §1.2

[18] The Independence Principle—as stated above—is vague in multiple ways; see (Christensen, 2019) for Christensen's latest revisions of and thoughts about the principle.

[19] See (Constantin and Grundmann, 2020; Grundmann, 2019; Zagzebski, 2012) for further details about "bracketing" and so-called "Preemption Views".

Spinelessness

Let's next turn to some objections to conciliationism.

Conciliatory views have often been met with the objection that they lead to a certain form of skepticism, also known as "spinelessness", see for example (Besong, 2014; Elga, 2007). The idea driving this objection is that conciliatory views will have the unwanted consequence that we ought to give up on our treasured beliefs in the face of disagreement way too frequently. It seems fair to suggest that we disagree with epistemic peers—or even epistemic superiors—on many controversial topics in religion, politics, ethics, natural science etc. So, should we really adhere to accounts of disagreement where it's often implied that we ought to give up on our own views, or at least reduce our confidence in them?

In response to the objection, Elga (2007) has suggested that in real-life cases of disagreement we very frequently happen to have dispute-independent reasons to downgrade the epistemic credentials of our interlocutors, and thus we'll in fact be able to avoid the problem of spinelessness on most occasions.

One reason in favor of this response is that in actual cases of moral and/or political disagreement, for example, we tend to disagree not just about a single proposition, but a wide range of interrelated issues—i.e., what Kappel and Andersen call 'comprehensive moral disagreement' (2019, p.1112):

> **Comprehensive Disagreement**. [Subject] A believes that p, where this belief is part of A's much wider web of moral beliefs. B disagrees about p, but also disagrees with many [other] parts of A's web of moral beliefs.

In order to grasp this kind of disagreement, think for instance of disputes between stereotypical conservatives and liberals in the US. Typically, such disputes are not just over an insulated issue, but a cluster of intertwined views about abortion, gay rights, gender equality, environmental protection, nuclear energy etc. This is well-supported by empirical results such as (Kahan and Braman, 2006). If a liberal knows a conservative's position on abortion, they will probably also know the conservative's position on gun control, global warming, and many other issues (at least in an American context).

Note, finally, that some have seen the spinelessness-objection as completely misguided. Instead of charging the conciliationist with an unhealthy skeptical attitude towards disagreement, one could simply bite the bullet and see conciliatory views as being intellectually humble, and for that very reason epistemically virtuous (Feldman, 2006; Christensen and Lackey, 2013; Matheson, 2015).

Self-Defeat

Another objection which is frequently leveled against conciliatory views is that they seem self-undermining in certain ways (Elga, 2010; Christensen and Lackey, 2013; Decker, 2014; Mulligan, 2015; Littlejohn, 2020; Knoks, 2022). Elga (2010) puts one version of the challenge as follows:

> Just as people disagree about politics and the weather, so too people disagree about the right response to disagreement. For example, people disagree about whether a conciliatory view on disagreement is right. So a view on disagreement should offer advice on how to respond to disagreement about disagreement. But conciliatory views on disagreement run into trouble in offering such advice.
>
> The trouble is this: in many situations involving disagreement about disagreement, conciliatory views call for their own rejection. But it is incoherent for a view on disagreement to call for its own rejection. So conciliatory views on disagreement are incoherent. That is the argument. (Elga, 2010, pp. 178-179)

According to Elga, the argument expressed in this quote is a clear knock-out of conciliationism. In a lovely one-line paragraph entitled '*Reply to the Self-Undermining Problem for Conciliatory Views*' he writes:

> There is no good reply. Conciliatory views stand refuted. (Elga, 2010, p. 182)

Whether Elga is right about this, we'll leave as an open question.

2.2 Steadfast Views, Rational Uniqueness, and Epistemic Justification

Steadfast views aim to avoid the conclusion that peer disagreements always require us to conciliate. According to such views, it's sometimes not the case that all parties of a peer disagreement ought to reduce their initial confidence with respect to the proposition under dispute. Thus, all steadfast views pay careful attention to *symmetry breakers*, i.e., facts that will allow at least one party of the disagreement in question to assign more weight to their own epistemic position.

Kelly's Right Reasons View

Tom Kelly (2005) has defended a radical steadfast view known as the *Right Reasons View* ('RRV').[20] Kelly argues that if one has epistemic justification for target-proposition $\langle p \rangle$—prior to peer disagreement—then the higher-order evidence one gets from the disagreement itself won't suffice to defeat one's initial justification.[21] Thus, according to RRV, if one has epistemic justification with respect to $\langle p \rangle$, prior to disagreeing with a peer about it, one is not rationally obliged to reduce one's confidence in $\langle p \rangle$ in the face of disagreement.

Let's look at Kelly's view in a bit more detail. Suppose subject S believes that $\langle p \rangle$ on the basis of first-order evidence E. Assume that S now discovers a disagreement as to whether $\langle p \rangle$ with epistemic peer S^*. From this fact of disagreement S doesn't get any first-order evidence suggesting that target-proposition $\langle p \rangle$ is false, Kelly says, S merely gets higher-order evidence suggesting that E doesn't support $\langle p \rangle$ after all; and according to RRV, S's justification for believing that $\langle p \rangle$ depends solely on whether E in fact supports $\langle p \rangle$, not on whether S's belief that E supports $\langle p \rangle$ is justified or not. Naturally, if it turns out that E in fact doesn't support $\langle p \rangle$, S should reduce confidence with respect to $\langle p \rangle$. Yet, this natural result is *not* explained by the epistemic significance of peer disagreement, but rather by the fact that this is what S's first-order evidence supports.

Kelly (2005) offers multiple arguments against the idea that the higher-order evidence constituted by peer disagreement defeats your initial doxastic attitude toward the target-proposition in question. One such argument is that, according to him, we typically don't refer to higher-order evidence when providing reasons for beliefs. Kelly asserts that, when one presents one's reasons for believing that $\langle p \rangle$ in mundane circumstances, one would typically point to the relevant first-order evidence E that bears on $\langle p \rangle$ rather than pieces of higher-order evidence like, say, the fact that one has inferred $\langle p \rangle$ from E.

This argument is hardly impressive. In fact, it's a bit hard to even see the relevance of Kelly's point. For all we know, what people typically do when they give reasons, form beliefs etc., might be completely off-track, unjustified, and irrational! The massive empirical literature on *heuristics and biases* à la Kahneman & Tversky (Kahneman et al., 1982; Tversky and Kahneman, 1981) would indeed suggest that humans are severely irrational across many ordinary

[20] Michael G. Titelbaum (2015) has presented a very interesting argument in favor of RRV that involves his so-called "Fixed Point Thesis". The reader should consult Appendix 1 of the present monograph for a thorough presentation of Titelbaum's intricate argument. More recently, Skipper (2019) has written about the connection between higher-order defeat and RRV.

[21] Note that epistemic justification comes in various forms and this might complicate Kelly's view a lot if taken seriously. See for instance (Littlejohn, 2012) for a tripartite division of epistemic justification into *propositional*; *doxastic*; and *personal*. Another distinction which might be relevant in the context of RRV is between *evidential* and *truth promoting non-evidential reasons* for belief, see (Talbot, 2014). See also (Conee, 1987) for a similar distinction, and note finally the discussions of the synchronic/diachronic-distinction in Bayesian epistemology (Talbott, 2016).

contexts.

Another argument considered by Kelly (2005) is that the higher-order evidence generated by peer disagreement seems to cancel itself out. Suppose that prior to the discovery of disagreement, you believe that $\langle p \rangle$ based on E. When discovering disagreement, however, two additional pieces of evidence become available to you, viz., the fact that you have inferred $\langle p \rangle$ from E and the fact that your peer hasn't. Kelly claims that the additional evidence owing to the discovery of disagreement doesn't call for a change in one's doxastic attitude toward $\langle p \rangle$. Because—assuming that equal weight is assigned to the higher-order evidence concerning yourself and to the higher-order evidence concerning your peer—we can reasonably infer that these weights will cancel each other out. This would in effect leave you with just the initial evidence E to rely upon.

In response to this, it has been suggested that the argument from "cancelling out" seems to miss the point that in many realistic cases one has extensive higher-order evidence about oneself and one's relevant inferential behavior *prior* to discovering peer disagreement, e.g., track record-evidence concerning one's previous performances in various domains as well as one's general qualifications and attitudes with respect to the central question at hand. If your doxastic attitude toward $\langle p \rangle$ already reflects such higher-order evidence about yourself prior to the encounter with an epistemic peer—crucially including the fact that you are epistemically qualified to judge whether $\langle p \rangle$ is true and that you have in fact inferred $\langle p \rangle$ from the original body of evidence E—then the new discovery of disagreement can change what you are justified in believing with respect to $\langle p \rangle$ post-disagreement (or at least this is what Matheson (2015, pp. 39-41) claims).

First-Person Views

Consider next a prominent class of steadfast views, which we call '*First-Person Views*'. According to such views we are entitled to accord our own basic intuitions (and sometimes considered judgements) with greater epistemic weight than those of others simply because they are our own (Enoch, 2010; Plantinga, 2000; Wedgwood, 2010).

In line with this, Foley (2001) has suggested that *self-trust* entitles us to discount the epistemic position of epistemic peers in disagreement scenarios, which looks like a brute rejection of the Independence Principle (cf. §2.1).

Similarly, Wedgewood (2007) has argued that the symmetry in cases of peer disagreement is an illusion. For if one takes seriously the significance of a first-person perspective *vis-à-vis* disagreement, then one is entitled to a higher degree of trust in one's own cognitive faculties than in the faculties of others.

The obvious problem facing first-person views like these is that of motivating why one's own perspective is epistemically privileged. In peer disagreement, the possibility that I have made an error and that you haven't seems to be a salient one, which *prima facie* calls for epistemic modesty. Hence, it seems overly dogmatic to reject this possibility out of pure self-trust. After all, what reason do I have to suppose that my basic seemings are more accurate than yours?

Arguments from the Falsity of Uniqueness

Another way of motivating steadfast views we'll consider goes via a denial of the *Uniqueness Thesis* (also known as *Rational Uniqueness*). The Uniqueness Thesis ('UT') concerns a relation between a body of evidence, a doxastic attitude, and a proposition. Jonathan Matheson—a proponent of the thesis—defines UT as follows:

> For any body of evidence E and proposition $[p]$, E justifies at most one doxastic attitude toward $[p]$. (Matheson, 2011, p. 360)

UT features frequently in the epistemology literature (Conee, 2010; Rosa, 2012; Kelly, 2014; White, 2014; Kopec and Titelbaum, 2016; Kauss, 2023), especially in debates concerning the possibility of rational peer disagreement: If two epistemic peers disagree as to whether $\langle p \rangle$, is it then possible for both of them to be justified in their doxastic attitudes toward $\langle p \rangle$? If UT is true, the answer is negative.

Importantly, there are several non-equivalent definitions of UT in the literature. Kelly (2010), for example, favors a formulation of UT stating that there is *exactly one* justified doxastic attitude given a body of evidence, while Matheson prefers an *at most one*-formulation, as we have just seen. Matheson notes that in most cases there will be *exactly one* justified doxastic attitude given a body of evidence, but in some situations, there may be no justified doxastic attitude toward $\langle p \rangle$ whatsoever. This can arguably happen when one is not able to, or when it is simply not possible to, comprehend the proposition at hand.[22] If one takes (possible) comprehension of $\langle p \rangle$ to be a necessary condition for the existence of a justified doxastic attitude toward $\langle p \rangle$, it seems most reasonable to use Matheson's weaker definition of UT. Thus, this is what we'll assume in the following.

Further we'll adopt Matheson's assumption that the term 'doxastic attitude' can only refer to the following three possibilities: *belief that* $\langle p \rangle$; *disbelief that* $\langle p \rangle$; and *suspension of judgement with respect to* $\langle p \rangle$. That is, the possibility space of doxastic attitudes that one can take toward a given proposition $\langle p \rangle$ is exhausted by these three (cf. the Tripartite View from §1).

So, UT puts a constraint on the total number of doxastic attitudes that a body of evidence can justify toward a proposition. According to UT any body of evidence E justifies at most

[22] See (Feldman, 2006) for a motivation of this view.

one doxastic attitude toward $\langle p \rangle$. In other words, there exists no body of evidence E such that E justifies both belief and disbelief toward $\langle p \rangle$. And similarly, of course, UT implies that there exists no E such that E justifies both a (dis)belief in $\langle p \rangle$ and suspension of judgement with respect to $\langle p \rangle$.

In his paper entitled '*The Case for Rational Uniqueness*', Matheson makes two further clarifying remarks about UT:

> (UT)...makes no reference to individuals or times since (UT) claims (in part) that who possesses the body of evidence, as well as when it is possessed, makes no difference regarding which doxastic attitude is justified (if any) toward any particular proposition by that body of evidence.[23] (Matheson, 2011, p. 360)
>
> (UT) concerns propositional justification, rather than doxastic justification. That is, the kind of justification relevant to (UT) is solely a relation between a body of evidence, a doxastic attitude, and a proposition. How individuals have come to have the doxastic attitudes they have toward the proposition in question will not be relevant to our discussion. Further, individuals can be propositionally justified in adopting attitudes toward propositions which they psychologically cannot adopt...Importantly, it is not a necessary condition for being justified in believing [p] that one be able to demonstrate that one is justified in believing. (Matheson, 2011, pp. 360-361)

The first of these quotes states that according to UT a given body of evidence E justifies exactly the same doxastic attitude (if any) towards $\langle p \rangle$, no matter the subject that assesses E and at what time this is done.

In the second quote, Matheson distinguishes between *propositional* and *doxastic* justification, where the former concerns a relation between a body of evidence, a doxastic attitude, and a proposition, the latter concerns *how* a given individual came to adopt a specific doxastic attitude toward a proposition, i.e., doxastic justification is concerned with one's reasons for adopting a certain attitude toward $\langle p \rangle$. Doxastic justification presumes that a given individual has a certain attitude toward $\langle p \rangle$, and the question is then whether or not this individual

[23] Note that while Matheson's statement of UT doesn't make reference to individuals (i.e., cognizers or human agents) at all, some authors have actually presented versions of uniqueness that do. Consider for example Titelbaum and Kopec's tripartite distinction between propositional, attitudinal, and personal uniqueness (Titelbaum and Kopec, 2019, p. 206).

Propositional Uniqueness. Given any body of evidence and proposition, the evidence all-things-considered justifies either the proposition, its negation, or neither.

Attitudinal Uniqueness. Given any body of evidence and proposition, the evidence all-things considered justifies at most one of the following attitudes toward the proposition: belief, disbelief, or suspension.

Personal Uniqueness. Given any body of evidence and proposition, there is at most one doxastic attitude that any agent with that total evidence is rationally permitted to take toward the proposition.

has sufficiently good epistemic reasons to be justified in having that attitude. When it comes to propositional justification, on the other hand, it is irrelevant whether any individual is ever concerned with $\langle p \rangle$; the crux of propositional justification is that a justification-relation between a body of evidence, a doxastic attitude, and a proposition holds, not whether any individual realizes this. An individual can thus be propositionally justified in a doxastic attitude towards $\langle p \rangle$ even though this individual has not adopted the relevant attitude psychologically. And hence, it's not necessary for a subject to be able to demonstrate or defend a given attitude toward $\langle p \rangle$ in order for it to be propositionally justified. Matheson tells us that UT is a thesis concerning propositional justification rather than doxastic justification.[24]

Now, how can one argue for steadfastness via a denial of UT? The answer is both straightforward and indicated above. If UT is false then it can be rational to adopt different (and incompatible) doxastic attitudes toward $\langle p \rangle$ based on the same evidence. So, when considering the evidence which is generated by a peer disagreement it need not be evidence to the effect that either of us is irrational. Hence, there is no immediate reason for any of us to adjust our initial doxastic attitudes upon discovering our peer disagreement (if we assume that UT is false) (Rosen, 2001).

We'll not discuss the plausibility of UT much further here, but simply point out that—upon reflection—the falsity of UT doesn't do a lot of work in the favor of steadfastness.[25] To see why, consider *Extreme Permissivism about Evidence*, i.e., a view according to which any doxastic attitude toward proposition $\langle p \rangle$ is justified by evidence E, where E is arbitrary. On this view, disagreements are necessarily unable to provide evidence to the effect that one's doxastic attitude is unjustified. (Well, perhaps disagreements wouldn't even be possible given an endorsement of this extreme view). But such an extreme form of permissivism is absurd and thus we are pushed towards a more moderate account immediately. As soon as we adopt a more moderate permissivist stance, however, there is room for acknowledging the epistemic significance of peer disagreement. For, suppose that evidence E is permissive in the sense that any credence within the interval $[0.6, 0.9]$ is rational to adopt *vis-à-vis* proposition $\langle p \rangle$. If so, peers S and S^* in disagreement as to whether $\langle p \rangle$ are each running an equal risk of adopting an irrational credence which lies outside the interval. Hence, it seems that both parties should conciliate if anybody should.[26] [27]

[24] Note here again Littlejohn's tripartite division of epistemic justification which includes *personal* justification as well as doxastic and propositional, see for instance (Littlejohn, 2012, p. 5). According to LittleJohn, doxastic justification is sufficient for personal justification, but not vice versa.

[25] See (Ross, 2021) for a recent discussion of the plausibility of UT and of some alleged counterexamples to the thesis.

[26] Thanks to Bjørn Gunnar Hallsson for providing the original argumentation in this paragraph; and generally for the many fruitful discussions about UT, peer disagreement, and rational credences, we have had over the years.

[27] Note also the fascinating work on mushy credences, which might be relevant here. See for example (Fraser, 2022).

Hybrid Views

Before ending our run-through of steadfast views, it's worth highlighting an important subclass of views which we'll call 'hybrid' in lack of a better name. These views will trivially count as steadfast because they deny the universality claim of conciliationism, and yet they stand out due to their distinctive *case-by-case* approach to peer disagreement. Below we'll consider Kelly's *Total Evidence View* as well as Lackey's *Justificationist View*.[28] Finally, we'll briefly mention a *knowledge-first* view due to Srinivasan and Hawthorne.

In (2010) Kelly defends a *Total Evidence View* ('TEV'), which replaces his earlier Right Reasons View. Kelly's main motivation for adopting TEV is a specific worry he has concerning standard conciliationism. According to him, conciliatory views run the risk of having the epistemic influence of first-order evidence "swamped" by the higher-order evidence arising in peer disagreement cases. If one follows TEV, on the other hand, the appropriate response to such disagreements will be a case-by-case matter, and a function of one's *total evidence* (as well as one's response to the evidence). That is, according to TEV, if S's response to the first-order evidence—prior to peer disagreement—is perfectly rational, the epistemic significance of the higher-order evidence (constituted by the disagreement itself) will be less than had S responded irrationally to begin with. Suppose, for instance, that S has responded properly to first-order evidence E while S^* hasn't, prior to peer disagreement; then S's belief that $\langle$*S has responded properly to E*$\rangle$ will be more justified[29] than S^*'s ditto towards $\langle$*S^* has responded properly to E*$\rangle$ upon realizing peer disagreement.

Of course, Kelly isn't suggesting that one is always in position to tell whether one has *in fact* responded in a rational way to the first-order evidence—that would be too unrealistic. Rather he's willing to accept a significant amount of intransparency as an epistemic deficit of the cognizers we are interested in, i.e., the intransparency as to whether one is responding rationally to a given body of evidence is simply an unfortunate upshot of the general human condition, and thus something that proponents of TEV will have to live with (along with everyone else).[30]

Note that Kappel (2019a) has argued against Kelly's TEV. Specifically, Kappel objects to what he calls the 'upward epistemic push' posited by TEV. According to Kappel, it's not the case

[28] According to the Total Evidence View from (Kelly, 2010) it's the interaction between the first-order and higher-order evidence, and one's response to the *total* evidence, that decides what is epistemically rational in cases of peer disagreement. In contrast, Kelly's earlier account—i.e., the Right Reasons View (2005)—took the first-order evidence to be dominant in determining what is rational in peer disagreement. One reason to subsume Kelly's Total Evidence View under the heading 'hybrid' rather than simply 'steadfast' is that the correct assessment of peer disagreement will be a case-by-case matter on this view. So, while Kelly still favors steadfastness in 2010, it is in a hybrid-version which is more open to the influence of higher-order evidence.

[29] See (Hawthorne and Logins, 2021; Fassio and Logins, 2023) for some interesting recent work on the gradability of epistemic justification.

[30] See also the important general discussion of *anti-lumniousity* in (Williamson, 2000).

that epistemic justification at the first-order level impacts on epistemic justification at higher-order level; the justificational impact runs in the opposite direction only. To illustrate Kappel's objection, consider the following. Suppose S's justification for believing the higher-order proposition that $\langle$*S has responded rationally to first-order evidence E for p*$\rangle$ is defeated in some way, then consequently S's justification for believing the target-proposition $\langle p \rangle$ at first-order level can be defeated as a result of the *downward epistemic push* stemming from the defeat of the belief regarding the higher-order proposition (and this can even be true in cases where S has in fact responded rationally to the first-order evidence). However, Kappel isn't convinced that there is an epistemic push going in the opposite direction, i.e., from first-order to higher-order level.[31]

In the present introductory chapter, we'll not consider the intricate details of Kelly's main argument in favor of the upward epistemic push, viz., the Argument from Recognition (2010), nor the details of Kappel's counterargument (2019a); instead we'll simply observe that the plausibility of TEV rests on controversial issues regarding the interconnections between first-order and higher-order levels of evidence, defeaters, and epistemic justification.

Next, let's turn to the hybrid view that Jennifer Lackey (2010) has defended. Since Lackey's view is *Justificationist* in spirit, we call it 'JV'. JV claims that no doxastic revision is epistemically required in peer disagreement exactly when your belief that $\langle p \rangle$ is highly justified and you have a relevant symmetry breaker (regarding your peer). In contrast, a drastic doxastic revision is required in peer disagreement exactly when your belief that $\langle p \rangle$ is sparsely justified and you don't have a symmetry breaker of the relevant kind. Moderate revision is called for in all intermediate cases.

What is particularly interesting about JV for our present purposes is that it combines classically *internalist* and *externalist* features from the literature on epistemic justification (Pappas, 2023). On the one hand, Lackey submits that symmetry breakers can stem from purely *introspective information*, which your peer cannot access. On the other hand, the notion of justification that Lackey adheres to isn't a purely internalist one as it relies on a necessary

[31] An illustrative case of *higher-order defeat* is given by Skipper (2019, p. 1373):

> **Parental Bias**. Mary rationally believes that her son Peter is a brilliant pianist. This morning, however, Mary reads a study showing that most parents suffer from a pronounced parental bias, which leads them to overestimate their children on a wide range of desirable traits such as intelligence, musical talent, social skills, and the like.

By assumption, Mary's initial evaluation of Peter's abilities on the piano is rational (unbiased), but still it seems that she ought to lower confidence that her son, Peter, is a brilliant pianist after learning about the parental bias. For even if Mary doesn't *in fact* suffer from the parental bias, she seems to have sufficient reasons to think that she does. So, perhaps Mary's epistemic situation requires her to give up her belief (or lower her confidence) that her son Peter is a brilliant pianist. Recent literature on higher-order evidence contains a host of similar cases in which some fully rational agent seems required to change her doxastic state, because she gets misleading higher-order evidence indicating that her current state is rationally flawed.

reliabilist constraint.[32] This suggests that Lackey's view isn't just hybrid due to its case-by-case approach to the epistemology of disagreement, but also because of its amalgamation of internalism and externalism. To illustrate the merits of JV, Lackey compares the original **Restaurant** with:

> **Extreme Restaurant**. While dining with four of my friends, we all agree to leave a 20% tip and to evenly split the cost of the bill. My friend, Mia, and I rightly regard one another as peers where calculations are concerned—we frequently dine together and consistently arrive at the same figure when dividing up the amount owed. After the bill arrives and we each have a clear look at it, I assert with confidence that I have carefully calculated in my head that we each owe $43. In response, Mia asserts with the same degree of confidence that she has carefully calculated in her head that we each owe $450, which is more than the total cost of the bill. (Lackey, 2010, p. 321)

According to JV, I'm required to revise my confidence in **Restaurant** due to the epistemic significance of peer disagreement, but not in **Extreme Restaurant**. Since in **Extreme Restaurant** my confidence that the share is $43 is highly justified due to the first-order evidence and my track record in the relevant domain. Further, I hold personal information that can act as a symmetry breaker. I know that I am being sincere, that I have slept well, that I'm not drunk, and so on. But given Mia's extreme response, I have no similar knowledge about her. For all I know, Mia might be insincere, sleep deprived, and intoxicated. This asymmetry in personal information conjoined with my high degree of justifiedness, means that I shouldn't revise my initial doxastic attitude even though we disagree.

Finally, let's turn to a knowledge-first view of peer disagreement before ending the section.[33] In (2013) Srinivasan and Hawthorne assert that any hope of offering general answers to the Doxastic Disagreement Question (cf. §1) is in vain. Instead they offer a *Knowledge Disagreement Norm* ('KDN'):

> *KDN*. In disagreement as to whether $\langle p \rangle$ (where S believes that $\langle p \rangle$ while S^* believes not-$\langle p \rangle$):

1. S ought to trust S^* and believe that not-$\langle p \rangle$ if and only if were S to trust S^*, this would result in S's knowing not-$\langle p \rangle$;

[32] In rough outline *Process Reliabilism* (i.e., the most common kind of reliabilism) is the following view: a belief-token b is epistemically justified if and only if b is caused/sustained by a reliable process. Here, a *reliable process* is taken to be a process of belief formation that (would) produce(s) a sufficiently high ratio of true to false beliefs given a specified set of circumstances and a domain of application. Process Reliabilism was first proposed and defended by Alvin Goldman. See for example (Goldman, 1979, 1986).

[33] The canonical work in *knowledge-first epistemology* is (Williamson, 2000).

2. S ought to dismiss S^* and continue to believe that $\langle p \rangle$ if and only if were S to stick to her guns this would result in S's knowing that $\langle p \rangle$, and;

3. in all other cases, S ought to suspend judgement about $\langle p \rangle$.

Notice how this view also presupposes the Tripartite View of doxastic attitudes with which we are now familiar (cf. §1). Note further that although KDN is a *knowledge*-centric norm, this is inessential to its possible significance. One could just as well have proposed similar norms like "TDN" or "JDN", in case one preferred a *truth-* or *justification*-centric epistemology.

What might be more worrying about KDN, is that one will not always be in a position to tell what exactly KDN recommends in concrete scenarios. That is to say, some cases of disagreement will be intransparent to the agents involved, e.g., cases where one knows that $\langle p \rangle$ but fails to know that one knows that $\langle p \rangle$, or cases where one doesn't know $\langle p \rangle$ but isn't in a position to know this.[34] In such cases, even if one knows that one ought to conform to KDN, one is not in a position to know what specific alternative of (1)-(3) to adopt in order to achieve KDN-conformity. In other words, KDN is not perfectly operationalizable.[35]

We'll not discuss the plausibility of KDN any further here, but instead make do with what Srinivasan and Hawthorne take to be the upshot of their knowledge-first approach to the epistemology of disagreement:

> We have suggested that those of us who hope for a general and intuitively satisfying answer to the question that is at the centre of the disagreement debate—namely, what we ought to do, epistemically speaking, when faced with disagreement—might be hoping in vain. There are deep structural reasons why such an answer has proven, and will continue to prove, elusive. Intuitively, we expect epistemic norms to be normatively satisfying: that is, we expect them to track our intuitions about blameworthy and praiseworthy epistemic conduct. An epistemic norm that ties what one ought to do to a non-transparent condition (e.g. knowledge) is an epistemic norm that will not satisfy this basic desideratum. To construct an epistemic norm that is normatively satisfying, then, we require an epistemic 'ought' that is tied to only transparent conditions; unfortunately, no such conditions plausibly exist. As such, the hope of finding a norma-

[34] See (Williamson, 2000) for concrete arguments against the so-called 'KK-Principle' (also known as 'Positive Introspection').

[35] Say that a norm N is *perfectly operationalizable* if and only if whenever one knows N, and is in circumstances C, one is in a position to reason as follows: (1) I am in C, (2) I ought to X in C, (3) I can X by φ-ing (where 'φ' refers to a basic mental or physical action-type that one knows how to perform). Clearly, KDN doesn't satisfy perfect operationalizability since knowledge-related conditions are intransparent (or "anti-luminous"), i.e., one is not always in a position to know whether C obtains.

> tively satisfying answer to the disagreement question seems like a hope unlikely to be satisfied (Srinivasan and Hawthorne, 2013, p. 28).[36]

While this message might seem bleak, we should grant that Srinivasan and Hawthorne are on to something in stating it. As we'll see in Chapter 7, several central principles from the peer disagreement-literature are at best deflated in the context of *logical* disagreement. So the hope of finding a sweeping and normatively satisfying answer to the Doxastic Disagreement Question is *in fact* in vain. That will be one of the main conclusions of the present monograph.

[36] See (Broncano-Berrocal and Simion, 2021) for a more recent knowledge-first view in the peer disagreement debate.

Chapter 1

First Preamble

1 Preamble

As is the case with all three preamble chapters of this monograph, the present one aims to provide some useful context for the research chapter that immediately succeeds it. On the following pages the reader will find an intuitive guide to the Ad Hoc Reading of 'logical disagreement' along with some important contextualizing information from the philosophy of logic and mainstream epistemology (§§1.1-1.4).

1.1 'Logical Disagreement'—The Ad Hoc Reading

The easiest way to introduce the Ad Hoc Reading of 'logical disagreement' is by contrasting it to the Theory Choice Reading. So we begin our preamble with a short detour.

Inspired by the philosophy of science—where *theory choice* is an established topic (Reiss and Sprenger, 2020)—the fashionable reading of 'logical disagreement' in contemporary philosophy of logic is the Theory Choice Reading (which will be our guiding thread through Chapters 3 and 4). In a nutshell this interpretation says that genuine logical disagreements take place when entire logical theories come into conflict with each other; as opposed to disagreeing about sub-theoretic claims in a piecemeal fashion. On the Theory Choice Reading, logical disagreements concern how we justify our choice of a whole logical theory such as classical, intuitionistic, relevantist, connexive, quantum etc., rather than what we believe about a particular logical principle or inference, say, Disjunctive Syllogism.[1]

[1] To be sure, the term 'logical theory' must at minimum be understood as a set of sentences logically closed under a given entailment-relation, and according to Ole Hjortland there is something like a consensus that the

Illustrative examples of logical disagreement at the level of logical theories are easy to find when one considers the plethora of disputes between classical and non-classical logicians in the history of philosophy. In such cases we often see that each side of the dispute tries to show that their favored logical theory captures the "one true logic"; or in modern abductivist terms, that their theory fits better with the relevant data (frequently taken to be our intuitive judgments about logical inferences).

To give a paradigmatic example, consider the proposition expressed by the sentence *No contradictions are true*, which seems to be an uncontroversial logical truth from the perspective of the classical logician. Many working scientists would even say that it provides the very mechanism through which one can demonstrate the falsity of theories in general—using reductio and/or empirical evidence (Martin, 2021c). Still the status of the proposition ⟨*No contradictions are true*⟩ has come under sustained attack from dialetheism, i.e., the view that some contradictions are true, in recent times (Priest, 2005, p. 1). The most persistent motivation for dialetheism is its modern dialetheic solutions to self-referential paradoxes such as the Liar Paradox, Russell's Paradox, Curry's Paradox etc. (Priest, 2006; Beall, 2011). According to dialetheists, these paradoxes have evaded successful non-dialetheic solutions due to an inherent flaw that all non-dialetheic solutions share rather than a lack of scrutiny amongst logicians. So, while the classical logician might claim that their classical theory is the best one because of its indispensability in scientific practice and discourse, a dialetheist can dispute the top position of classical logic by appealing to ingenious dialetheic solutions to paradoxes that have plagued the enterprise of logic for ages. As we shall see in Chapter 4, contemporary abductivists like Timothy Williamson (2017b) and Graham Priest (2014) hold that the grounds for choosing one logical theory over another is how well it fits with relevant data plus its theoretical virtues and lack of vices, e.g., its strength in terms of ratified consequences, how aesthetically elegant and simple it is, and how ontologically parsimonious.

Now—in contrast to the Theory Choice Reading of 'logical disagreement'—the Ad Hoc Reading is meant to capture logical disagreements not just at theory-level, but also piecemeal logical disputes taking place at a *sub-theoretic* level. Some logical disagreements under this interpretation will bear closer resemblance to mundane cases like **Restaurant** and **Horse Race** than to theory-loaded quarrels from the philosophy of science.

One example of an ad hoc logical disagreement could be a controversy over a particular instance of the law of the excluded middle, e.g., ⟨p or not-p⟩. Say that you believe that ⟨p or not-p⟩ is true, while I don't. Then you might try to convince me to believe ⟨p or not-p⟩ by suggesting that a denial of a disjunction requires denial of each of its disjuncts, and if one denies ⟨p⟩, one cannot simultaneously deny ⟨not-p⟩ without contradiction. This would

main function of a logical theory is to tell us which inferences are valid (Hjortland, 2019, p. 252). However, some authors add to this deflationary understanding a demand that theories should account for features like provability, truth-preservation, formality, and consistency, as well (Priest, 2005; Hjortland, 2017).

seemingly be a dispute about the meanings of the logical constants given by 'or' and 'not'.[2]

Some—e.g., Quine (1960; 1986)—have suspected that cases of disagreement akin to this one are not genuinely logical, but merely verbal. In a famous passage, Quine discussed a clash between a classical and a non-classical logician, where the non-classical side allegedly changes the meaning of negation:

> My view of the dialogue is that neither party knows what he is talking about. They think that they are talking about negation, '[$\neg$]', 'not'; but surely the notion ceased to be recognisable as negation when they took to regarding some conjunctions of the form '$p[\wedge]$ $[\neg]p$' as true, and stopped regarding such sentences as implying all others. Here, evidently, is the deviant logician's predicament: when he tries to deny the doctrine he only changes the subject. (Quine, 1986, p. 81)

In order to avoid subsuming merely verbal disputes under the heading of genuine logical disagreements, Hattiangadi (2018, p. 92) has suggested imposing the following constraint on logical disagreement:

> *Genuine Disagreement Constraint*. Any adequate account of logical disagreement must be such that, if agents S and S^* genuinely disagree, then the assignment of attitudes and contents to S and S^* must explain their disagreement.

Importantly the intended reading of this constraint doesn't require that the assignment of attitudes and contents must explain *why* the agents disagree, e.g., by appealing to the causal

[2] Much more could be said about what the entities of logic mean, or what logic is about (if anything), but in order to avoid too many substantial detours from our main topic we'll have to make do with the common distinction between *Representationalism* and *Inferrentialism* here. The basic idea of the former is that we are confronted with entities of different sorts and somehow make our words (or logical symbols) stand for them. Within this paradigm, the essential expressions of our languages are meaningful insofar as they represent, i.e., stand for something. In the context of logic, an important example is Frege's suggestion that whole declarative sentences of a logic represent binary truth values, which he took to be abstract entities, namely The True/The False (Frege, 1948, 1956).

An alternative to representationalism—viz. inferrentialism—was put forward by the later Wittgenstein (1969a; 2009), and is often caricatured with the dictum: "*Meaning is use*". Wittgenstein's proposal was that we should see the relation between an expression and its meaning in a similar way to how we conceive the relationship between a chess piece (pawn, king, queen etc.) and its role in the game of chess. By itself this idea wasn't new, but Wittgenstein's massive influence was able to bring the relationship between meaning and rules of language games into the centerstage of philosophical research. In the context of logic, Neil Tennant (2007, p. 1056) states that: "*An inferentialist theory of meaning holds that the meaning of a logical operator can be captured by suitably formulated rules of inference...*".

See also (Gentzen, 1936; Sellars, 1953; Prior, 1960; Prawitz, 1965, 2006; Dummett, 1991; Restall, 2022; Restall and Standefer, 2023; Restall, 202X) for more elaborate discussions of inferentialism and proof-theoretic semantics.

histories of the involved agents, rather the assignment must explain *what* their disagreement consists in—e.g., inconsistency between the conflicting attitudes.[3]

While we won't endorse the Genuine Disagreement Constraint explicitly in what follows, we will assume that genuine logical disagreements are possible and not merely cases of talking past one another (as the quote from Quine would suggest). In other words, we'll deny the *Meaning-Variance Thesis*, stating that:

> Classical and nonclassical logicians are not engaged in a substantive debate about the nature of logical laws, but are simply attaching different meanings to the same expressions. Once the parties are clear on what they mean by locutions such as 'and', 'not', 'valid', 'proof', the conversation can proceed with the dispute resolved (Hjortland, 2022, p. 2).

Hence we'll be assuming an idealization of logical disagreements close to that of full disclosure from the peer disagreement debate (cf. §1.2 of the Introduction).

Another example of an ad hoc disagreement—which is perhaps closer to the heart of what is *logical* about logical disagreements—would be a disagreement over $\langle \varphi \vee \neg\varphi \rangle$. The target of our dispute is now formalized and generalized. You believe that $\langle \varphi \vee \neg\varphi \rangle$ is a true proposition stating a general principle, while I deny it. That is, a dispute as to whether there are any genuine counterexamples to the schema of $\langle \varphi \vee \neg\varphi \rangle$.[4]

A third example, which falls right out of the above one, is a disagreement over $\langle \Gamma \vDash \varphi \vee \neg\varphi \rangle$, where the symbol '$\Gamma$' denotes a set of premises. (Alternative notation for the double turnstile-symbol '$\vDash$' could be '$\Vdash$', '$\Rightarrow$' etc.). Say that you believe that $\langle \Gamma \vDash \varphi \vee \neg\varphi \rangle$ holds, while I suspend judgement about this. The target of our dispute is then whether a given conclusion is logically entailed by a set of premises (Hattiangadi, 2018, p. 88).[5]

As should be clear at this stage, the Ad Hoc Reading is not meant to exclude logical disagreements involving theory choice in logic, the point is rather that the Ad Hoc Reading is more inclusive than the Theory Choice Reading as it *can* include both theory-level and sub-theoretic logical disagreements under its heading. Indeed, as soon as one uses a symbol like the double turnstile, i.e., '$\vDash$', it's easy (for some even habitual) to associate piecemeal

[3] For further discussion of (merely) verbal disputes, see for example (Chalmers, 2011; Hjortland, 2014; Jenkins, 2014; Cohnitz, 2020).

[4] Here, some might suggest that what counts as a genuine counterexample is *context-dependent* and that there are well-known logical disputes from the literature in which someone wants to push the extension of acceptable counterexamples beyond its normal bounds, e.g., when Gillian Russell argues in favor of logical nihilism via very controversial counterexamples to basic logical laws (Russell, 2018a).

[5] Note that our use of the double turnstile-symbol '$\vDash$' is meant to indicate that we think of this and other examples in semantic terms rather than proof-theoretic ones. But a friend of proof-theory could have used '$\vdash$', '$\succ$' etc. just as well.

sub-theoretic disputes with clashes between different schools of logic even if this is not fully explicit. As Florian Steinberger (2022) observes:

> No one, of course, ever disputed that a given classical argument form is valid relative to the notion of validity-in-classical logic, where as an intuitionistically valid argument is valid relative to the notion of validity-in-intuitionistic logic. The problem is that there is no neutral notion of validity one could appeal to that would enable one to make sense of logical disputes as genuine debates, which, arguably, they are. What is needed to capture the substantive nature of these disputes, therefore, is a workable non-partisan notion of validity, one that is not internal to any particular system of logic.

In Chapter 4 we'll see that there are good reasons for thinking that a few special entailment-claims are in fact epistemically justified independently of any particular system (or theory) of logic, but Steinberger's statement does highlight something important for us nonetheless. In many (or even most) real-life cases of logical disagreement the validity of the argument forms involved will be relative to *deep* theoretical commitments among the combatants (more on deep disagreement in §1.4 of the present preamble and Chapters 2 and 7).

One notorious example of a logical disagreement—where validity is very often explicitly relativized to specific traditions of logic—is the controversy over Double Negation Elimination, i.e., $\langle\neg\neg\varphi \vDash \varphi\rangle$. Assuming the Tripartite View of doxastic attitudes from §1 of the Introduction above, the proponent of classical logic believes that $\langle\neg\neg\varphi \vDash \varphi\rangle$ holds, while the intuitionist disbelieves it (or suspends judgement). Another famous example concerns the validity of Ex Falso Quodlibet (or the principle of explosion), i.e., $\langle\varphi \wedge \neg\varphi \vDash \psi\rangle$, where proponents of the classical tradition believes that from a contradiction anything follows, dialetheists don't. Yet another well-known case is that of Tertium Non Datur (or the law of the excluded middle), i.e., $\langle\vDash \varphi \vee \neg\varphi\rangle$, the classical logician believes the truth of $\langle\vDash \varphi \vee \neg\varphi\rangle$ whereas for example a proponent of the trivalent (strong) Kleene logic doesn't because there aren't any valid formulas in their theory. A final example is the dispute regarding Aristotle's Thesis, i.e., $\langle\vDash \neg(\neg\varphi \rightarrow \varphi)\rangle$, which is found invalid by the classical logician while it is validated by some proponents of connexive logic.[6]

1.2 The Normativity of Logic

Something that will become relevant to us in the forthcoming chapters is the *normative* status of logic.

[6] Notice that the last example of logical disagreement (involving Aristotle's Thesis) is special because it features a *contra-classical* logician. All the other cases discussed above concern disputes between *sub-classical* and classical logicians.

Intuitively, we take agents who fall short of the demands of logic to be rationally defective in some way, which, at first glance at least, suggests that logic is normative for reasoning.[7] Consider for example the following passage from Michael Titelbaum:

> Suppose Jane tells us (for some particular propositions p and q) that she believes it's not the case that either the negation of p or the negation of q is true. Then suppose that Jane tells us she also believes the negation of q. $\neg(\neg p \vee \neg q)$ is logically equivalent to $p \wedge q$, so Jane's beliefs are inconsistent. If this is all we know about Jane's beliefs, we will suspect that her overall state is rationally flawed (Titelbaum, 2015, p. 277).

Here, Titelbaum's assessment of Jane's overall doxastic state lends motivation to the following principle (Cohnitz and Estrada-González, 2019, p. 183):

> *Logical Consistency Principle.* S ought to avoid having logically inconsistent beliefs.

A principle that will strike many as a mere platitude, and similarly we might normally take for granted that: if S's beliefs jointly imply $\langle p \rangle$, then S ought to believe that $\langle p \rangle$ (cf. the *Logical Implication Principle* (Cohnitz and Estrada-González, 2019, p. 183)).

That logic has a normative role to play in our cognitive lives, underpinned by such principles, is indeed rooted in our traditional ways of thinking about logic. As is well known, Gottlob Frege classified logic as a "normative science" similar to ethics (Frege, 1997, p. 228) and in (2013, § 15) he wrote that:

> It is commonly granted that the logical laws are guidelines that thought should follow to arrive at the truth.

Similar remarks concerning the normativity of logic can be found in philosophical classics such as Kant's magnum opus *Critique of Pure Reason* (2003).

All this notwithstanding, Gilbert Harman has forcefully challenged the view that logic is normative for reasoning (Harman, 1984, 1986). Deductive logic and reasoning are two fundamentally different enterprises—logical principles are not in any direct sense rules of belief

[7] Strictly speaking it's too quick to simply assume that the demands of logic are normative for *reasoning* in particular (given that they are normative). Alternatively they could for instance be normative for governing *assertion*, as suggested by Milne in (2009), or for guiding certain multi-agent dialogical practices as proposed by Dutilh Novaes in (2015). See (Russell, 2020) for a recent argument suggesting that logic isn't normative in any distinguishing way. See (Arbeiter, 2023) for the novel idea that the concept of validity acts similarly to well-known thick concepts from ethics (Väyrynen, 2021).

revision, he argues (Harman, 1984, p. 107). Thus, Harman's work drives in a wedge between deductive logic and the norms of reasoning, and exposes the need for what John MacFarlane dubs 'bridge principles'. These principles are called for in order to illuminate the normative constraints that logic allegedly imposes on our reasoning (MacFarlane, 2004). The literature (Steinberger, 2019a; Evershed, 2021; Arbeiter, 2023) suggests that, in general, bridge principles are of the form:

> *Bridge*. If $\delta(\Gamma \vDash \varphi)$ then $D(\alpha(\Gamma), \beta(\varphi))$,

where δ is a doxastic attitude (judging, believing etc.) *vis-à-vis* the entailment $\Gamma \vDash \varphi$ (note that some bridge principles leave δ empty). D is a deontic operator (varying in type and scope) constraining the (possibly distinct) doxastic attitudes α, β *vis-à-vis* Γ, and φ, respectively.[8]

Some simple examples of bridge principles are:

- If $\langle \Gamma \vDash \varphi \rangle$, then subject S ought to believe $\langle \varphi \rangle$ if believing every member of Γ.
- If S believes that $\langle \Gamma \vDash \varphi \rangle$, then S ought *not* to believe every member of Γ and disbelieve $\langle \varphi \rangle$.
- If $\langle \Gamma \vDash \varphi \rangle$, then S has a *pro tanto* reason to believe $\langle \varphi \rangle$ if believing every member of Γ.

To get a better understanding of the different kinds of normative constraints deductive logic could impose on reasoning, let's follow Florian Steinberger (2019b,c) in distinguishing between directives, evaluations, and appraisals:

- *Directives.* First-personal instructions guiding doxastic conduct.
- *Evaluations.* Third-personal evaluative standards for classification of doxastic states as correct or incorrect.
- *Appraisals.* Third-personal norms underwriting attributions of blame and praise.

Clearly, one cannot expect deductive logic to deliver directives for human reasoners, i.e., first-personal guidance for our doxastic conduct. As fallible human agents we often have false

[8] To avoid making the formal notation of bridge principles any more clumsy, we simply take for granted that doxastic attitudes are to be had by cognitive agents (rather than indexing the symbols referring to doxastic attitudes to such agents). Note also that while we follow Steinberger (2019a, p. 312) in using the above formalism for generalized bridge principles, the notation is actually somewhat confusing, e.g., the operator D can vary in scope, but still it certainly looks as if it takes a wide scope in the formalism.

beliefs about what follows from what, and what doesn't follow, this is only expected given our cognitive limitations, and thus brute principles of deduction are quite unhelpful considered from a first-personal perspective of belief revision.

Also—as we saw in §2.2 of the Introduction—a norm N is said to be *perfectly operationalizable* if and only if whenever one knows N, and is in circumstances C, one is in a position to reason as follows: (1) I am in C, (2) I ought to X in C, (3) I can X by φ-ing (where 'φ' refers to a basic mental or physical action-type that one knows how to perform). But simple norms like Logical Consistency and Logical Implication don't satisfy perfect operationalizability since one is not always in a position to know whether C obtains. One can surely fail to believe consistently even if one knows one ought to; and similarly, one can fail to draw logical consequences that follow from one's beliefs even if one knows one should.

So logical principles don't deliver transparent and "followable" decision procedures for doxastic conduct, rather they provide criteria of rightness (Parfit, 1984).[9] We might know that inferring in accordance with Modus Tollens in circumstances C is the right thing to do from a third-personal perspective, but as the Wason Selection Task (1968) has shown, this doesn't mean that we are actually able to follow through in the heat of the moment.[10] The Wason Selection Task has been used in numerous experiments in cognitive psychology and other fields. In the Wason Selection Task, subjects are given four cards. In one version, each card has a number on one side, and a letter on the other, and subjects are given the cards facing 'A', 'K', '4', and '7' upwards for them to see. Subjects are then asked to decide which cards they must turn in order to assess the truth value of the proposition expressed by the statement: '*If there is a vowel on the one side of the card, then there is an even number of the opposite side*'. It's normally assumed that the correct solution is that one must turn the cards showing 'A' and '7' since only these two cards can disconfirm the statement. Turning the cards with 'K' and '4' yields no new evidence that one can use in determining the truth value of the proposition expressed by the statement. Interestingly, only about 10% of normal experimental subjects find the right solution, and this is widely believed to be a sign of irrationality, albeit, a kind of irrationality that most normal subjects happen to be prone to.

In line with this, it will frequently strike us as unfair to blame an agent for not realizing the truth of some deduction given that the agent lives up to all the third-personal norms of appraisal, say, typical epistemic virtues like curiosity, open-mindedness, thoroughness, intellectual humility etc. If the agent is doing everything that can reasonably be expected of them while just being in unfortunate epistemic circumstances, we might feel an urge to say that the agent is not only blameless, but also rational (or justified) in her doxastic affairs. This,

[9] In ethics, a *criterion of rightness* specifies the necessary and sufficient conditions for an action to be morally right (or permissible); whereas a *decision procedure* is some trait, disposition, method, rule, heuristic etc., that agents use more or less successfully for determining what action(s) they ought to perform in a given situation.

[10] Thanks to Bjørn Gunnar Hallsson and Klemens Kappel for many fruitful discussions about the Wason Selection Task and other studies in empirical psychology.

however, would be a mistake insofar as we take deductive logic to deliver evaluations, viz., third-personal standards of correctness. If the standards of deductive logic are mere evaluations, then it doesn't matter normatively speaking whether the agent is unlucky or not. All that matters is whether the agent satisfies the relevant standards. Accordingly, there can also be instances of "blameless wrongdoing" and "blameworthy right-making" in (logical) reasoning.

With the division between directives, evaluations, and appraisals, in mind we'll be able to see that the propositional justification discussed below in Chapter 2 will count as a third-personal evaluative standard for classification of doxastic states as correct or incorrect, i.e., an *evaluation* in Steinberger's terms. We'll also return to the importance of bridge principles in chapters 5 and 6 when discussing the epistemic (ir)rationality of logical akrasia.

1.3 Logical Evidence

Below—in Chapter 2—we'll see that the success of the *Argument from Logical Disagreement* depends on at least two controversial issues concerning evidence: (i) how should we individuate *logical* evidence in particular; and (ii) is evidence in general factive.

Let's first turn to (i). As suggested earlier (cf. §1.1), it is presently well received to conduct epistemological research about logic in a *holistic* fashion. In line with this Ben Martin has recently introduced a "practice-based" approach to the epistemology of logic (2022), which is inspired by the actual practices of working logicians as well as Quine's empirical holism (1953). Martin's epistemology explicitly opposes some traditional epistemologies of logic—viz., Rationalism and Semanticism—and asserts that logical propositions shouldn't be directly justified with intuitions and/or linguistic definitions as the evidential basis. Instead entire logical theories should be justified using a diverse pool of evidential sources, e.g., their ability to solve logico-semantic puzzles, accommodate the meaningfulness of natural-language sentences, and respect core practices of the mathematical sciences (Martin, 2021b).

Importantly, justification *holism* is not claiming that one cannot have justification for an individual (proposition expressed by the) claim that, say, '*Double negation elimination is valid*'.[11] For one could easily obtain such individual justification via a proof *within* some logical theory. The key point here is that, according to the holist, any such justification presupposes the context of an entire logical theory, and depends on a choice of such theory, e.g., choosing a classical theory rather than an intuitionistic one.

For the sake of Chapter 2, we'll individuate logical evidence *sub-theoretically*, which is in stark contrast to the holistic approach preferred by Martin and many of his contemporaries (e.g.,

[11] Let 'φ' denote a meta-variable and let the symbol '$\neg$' denote negation. Then double negation elimination is the entailment from $\neg\neg\varphi$ to φ.

Gill Russell (2019a)). When confronted with a deductive inference schema suggesting a doxastic transition from a set of premises Γ to a conclusion $\langle \varphi \rangle$, we'll simply think of Γ as (formal) first-order evidence in support of $\langle \varphi \rangle$ (even if the evidential relation isn't relativized to a particular logical theory). This way of carving things will perhaps seem natural to epistemologists from the mainstream tradition, including contemporary social epistemology, while it is bound to be problematic for those who prefer to think of logical disagreement in terms of entire logical theories competing with each other. According to the Theory Choice Reading, which is inspired by the philosophy of science-literature, it's simply unintelligible to individuate first-order logical evidence independently of theory choice—i.e., choosing a particular logical theory is a prerequisite for having any such evidence. No theory, no evidence!

Now, even though individuating logical evidence holistically might be thought of as the current default position in philosophy of logic, traditional views like *Rationalism* and *Semanticism* have individuated logical evidence sub-theoretically with some success. Here's a brief run-through of these traditional accounts.

In the most famous cases rationalists believe that we grasp the truths of logic (and other necessary truths) through a distinct form of intuition (BonJour, 1998). On a toy example of rationalism, having the intuition that proposition $\langle p \rangle$ is true is sufficient (though defeasible) evidence for having epistemic justification of the belief that $\langle p \rangle$. One famous proponent of logical rationalism was Kurt Gödel who believed that one can obtain knowledge about logic and mathematics qua direct intuition. With respect to set theory, he claimed that:

> [D]espite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as true. I don't see any reason why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception. (Gödel, 1964, p. 271)

Clearly any such rationalistic account owes us a plausible story as to why (mathematical) intuition provides genuine epistemic justification of logical propositions (and beliefs about them). In other words, it will have to explain why we can obtain epistemic goods in the domain of logic via our pure intellectual insight. As was the case with Gödel, many rationalists have simply taken our intuitions about logical propositions to provide defeasible evidence for logical truths because of their phenomenological resemblance to perceptual states, see for example (Bealer, 1998; Chudnoff, 2011; Koksvik, 2017).

Semanticism, on the other hand, tells a different tale. According to semanticism—often claimed by empiricists (subsuming the logical positivists)—propositions expressed by logical sentences are true solely in virtue of their meaning (Carnap, 2014). Thus, simply understanding or knowing the meaning of a given logical sentence is sufficient to gain evidence for (justifiably believing) the truth/falsity of the proposition it expresses. Exemplifying this view, Ayer writes:

> If one knows what is the function of the words 'either', 'or', and 'not', then one can see that any proposition of the form 'Either p is true or p is not true' is valid. (Ayer, 1952, p. 79)

One should appreciate that even though logical rationalism and semanticism are competing views in many ways—most notably with respect to the source of evidence and epistemic justification—they agree on the justification of logical propositions being purely *a priori*. While rationalism presumes some cognitive faculty, dedicated to the fostering of a priori-intuitions that can justify our (beliefs about) logical principles; semanticism claims that knowing the function of the meaning-constituting parts of logical sentences is sufficient for the justification of (beliefs regarding) such matters. This grasping of meaning might be counted as a priori since it concerns *relations of ideas* rather than *matters of fact* (Hume, 1975, Section 4). In other words, since logical truths are necessary (perhaps even self-evident) truths following linguistic definitions, they can plausibly be categorized as being independent of experience.[12]

Next—with respect to (ii)—though it's quite clear from Martin's holistic and practice-based approach to the epistemology of logic (Martin, 2021b, 2022; Martin and Hjortland, 2022, 202X) that he will take (logical) evidence to be non-factive, we'll do the exact opposite in Chapter 2. That is to say, we'll individuate logical evidence sub-theoretically and consider all evidence factive. Notice, however, that (i) and (ii) are orthogonal in the sense that while we happen to disagree with Martin regarding both (i) and (ii), it's perfectly possible to adopt a position where one only disagrees with him on one of the issues. In fact, this is what Tim Williamson does since he individuates logical evidence holistically—as Martin does—but *pace* Martin he takes evidence to be factive.

The three different positions *vis-à-vis* logical evidence just indicated are captured in the table **Logical Evidence** on page 36 below.

Further, it's interesting to observe that while it used to be the default position among mainstream epistemologists to consider evidence non-factive, a recent trend in the literature challenges this. Some have even suggested that epistemology is taking a regular "factive turn" nowadays (Mitova, 2018), supposedly originating with Williamson's Copernican turn away from the post-Gettier era and towards his knowledge-first programme (2000). Williamson notoriously equates evidence with knowledge in his principle $E = K$, i.e., he takes the true propositions that constitutes one's evidence to be coextensive with what one knows.

One way in which Williamson motivates $E = K$ is via formal models in epistemic logic of the Hintikka-tradition (Hintikka, 2005). According to Williamson, his factive view of evi-

[12] See (Williamson, 2007, Chapter 5) for various readings of the term 'a priori'.

Table 1.1: Logical Evidence ('A' is short for 'Andersen', 'M' for 'Martin', and 'W' for 'Williamson').

Logical Evidence$_A$	Factive	Non-Factive
Holistic	✓	×
Sub-Theoretic	✓	×
Logical Evidence$_M$	Factive	Non-Factive
Holistic	×	✓
Sub-Theoretic	×	×
Logical Evidence$_W$	Factive	Non-Factive
Holistic	✓	×
Sub-Theoretic	×	×

dence fits better with how logicians and formal epistemologists have thought about knowledge and epistemic accessibility (Williamson, 2000, 2011b, 2013b).

Following Williamson in (Skipper and Steglich-Petersen, 2019, Chapter 13), one can define an *epistemic frame* as an ordered pair $\langle W, R\rangle$, where W is a nonempty set of possible worlds that are mutually exclusive and jointly exhaustive, and R is a binary relation on W. In the present context, the individual points of the frame, viz., the various possible worlds, should merely be understood as relevantly specific—i.e., they need not be metaphysically possible in order to be *epistemically* accessible.

We can then model coarse-grained (i.e., not internally structured) propositions as subsets of W. For each subset, some specific proposition is true in every world of the set and false in every other. Thus, the subset relation will correspond to logical entailment, set-theoretic intersection to conjunction, union to disjunction, complementation in W to negation etc.

The relation R will specify the *total* evidence of a given agent (at a given time) such that $\langle w, x\rangle \in R$ if and only if it is consistent with one's total evidence in w that one is in x. Williamson defines $R(w)$ as $\{x : Rwx\}$, i.e., the strongest proposition to follow from one's evidence in w, and assumes that one's total evidence is consistent as a minimum requirement (on the pain of triviality).

Building on this, a *probabilistic frame* $\langle W, R, Pr\rangle$ is an ordered triple, where Pr is a probability distribution over W. The function Pr maps each subset of W to a real number between 0 and 1 such that $Pr(W) = 1$, and $Pr(X \cup Y) = Pr(X) + Pr(Y)$ whenever X and Y are disjoint. We assume that W must be *countable* and Pr *regular*, viz., $Pr(X) = 0$ only if $X = \emptyset$.

Informally speaking, Williamson regards Pr as a prior probability distribution. Posterior probabilities in a world, w, are defined by conditioning Pr on one's total evidence in w. So, the posterior probability of X in w, written '$Pr_w(X)$', is the prior probability of X conditional on $R(w)$:

(EVPROB). $Pr_w(X) = Pr(X \mid R(w)) = Pr(X \cap R(w))/Pr(R(w))$, where $Pr(R(w)) > 0$.[13]

The framework captures non-trivial propositions about probabilities on the evidence automatically, i.e., for any proposition X and real number c, we may define $Pr_{\geq c}[X]$ as $\{w : Pr_w(X) \geq c\}$, the proposition that the probability on one's evidence of X is at least c, which may be true in some worlds and false in others.

We shall not pursue Williamson's arguments for $E = K$ and the factivity of evidence any further here, but merely note that it's a bit peculiar—in the context of logical disagreement—how (parts of) his motivation for taking evidence as factive involve(s) significant chunks of classical logic (and set theory). Presumably, as a card-carrying holist, Williamson will need some evidence to support his choice of a classical theory of logic, and yet he motivates his view of evidence and evidential support by appeal to (modally extended) classical logic and the model-building it can be utilized for. While it's certainly plausible that "Williamson the Abductivist" is willing to bite the bullet here, a (vicious) circle lurks in the background nonetheless.

To be fair, this predicament of running in circles when trying to justify one's logical theory isn't uniquely Williamson's problem. Cashed out in more general terms the situation illustrated by Williamson's case is known as the '*Background Logic Problem*': In order to justify logical theory T on the basis of non-direct evidence E, it's required to make logical inferences regarding the consistency of E with T, and this will in turn presuppose the validity of certain rules of implication R (Wright, 1986; Shapiro, 2000; Martin, 2021b).[14]

1.4 Deep Disagreement

Influenced by the Background Logic Problem (and related issues such as Kripke's Adoption Problem, cf. Chapter 3) it is plausible to think that logical disagreements are cases of *deep disagreement*—which is a kind of disagreement we'll touch upon in Chapter 2 and then discuss more thoroughly in Chapter 7.

Frequently we disagree about trivial things such as where one finds the cheapest tofu in town or how tall a certain building is, but occasionally our disagreements run deeper. Sometimes we disagree about the very assumptions that facilitate our normal exchange of reasons and arguments. Recent epistemological parlance suggests that disagreements of the latter kind are "deep".

To get an intuitive grasp of deep disagreement consider the **Young Earth Creationist**:

[13] This definition is equivalent to **(BCOND)** from (Williamson, 2000, p. 214).

[14] Note also related writings on the so-called 'Logocentric Predicament'. See for instance (Ricketts, 1985).

> Henry is an Evangelical young Earth creationist, who accepts that the Earth is no more than 6000 years old and a nexus of conspiratorial claims as evidence of why scientists have been misleading us about the age of the Earth. Henry also rejects the theory of evolution and contemporary cosmology, citing literal readings of the Bible: 'your denial of scripture is unjustified', he says. Henry's neighbor Richard is a proponent of so-called 'New Atheism', and rejects the religious and young Earth creationist views of his neighbor Henry, and asserts that the Earth is much older than 6000 years: 'your denial of geology and evolutionary biology are unjustified', he says.[15] (Ranalli, 2021, p. 984)

This case has been widely discussed in the literature and is considered a paradigmatic case of deep disagreement (Lynch, 2010; Pritchard, 2010; Kappel, 2012; Hazlett, 2014; Ranalli, 2021; Ranalli and Lagewaard, 2022a,b).

Although there are several different ways of understanding the essentials of deep disagreement, we'll focus on the *Fundamental Epistemic Principle Theory* to avoid unnecessary detours.[16] According to this theoretical stance, deep disagreements are *deep* because they aren't solely concerned with "surface-level" propositions about, say, a particular weather forecast (Christensen, 2007), but also propositions stating the fundamental epistemic principles we ought to apply when trying to predict the weather in general. In other words, deep disagreements are disagreements over fundamental epistemic principles like those specifying which traditions, institutions, methods, sources of evidence, and patterns of reasoning to rely upon when forming beliefs (Kappel and Andersen, 2019).

Rational irresolvability is often considered a necessary property of deep disagreements because of their dialectical setup (Wittgenstein, 1969b; Fogelin, 2005; Lynch, 2010, 2016; Kappel, 2012). How is one supposed to give a compelling argument for target-proposition $\langle p \rangle$, when one's interlocutor asserts not-$\langle p \rangle$ (or suspends judgement as to whether $\langle p \rangle$), and does so by appealing to fundamental epistemic principles that conflict with one's own?[17] In the words of Michael Lynch:

[15] We'll craft a formal model of the deep disagreement scenario described in **Young Earth Creationist** in Chapter 7.

[16] According to Ranalli (2021), state of the art research on how to best characterize deep disagreement falls roughly into two theoretical camps. On the one hand we have the *Hinge Proposition Theorists* (Wittgenstein, 1969b; Feldman, 2005a; Fogelin, 2005; Friemann, 2005; Hazlett, 2014); on the other the *Fundamental Epistemic Principle Theorists* (Lynch, 2010; Kappel, 2012; Jønch-Clausen and Kappel, 2015; Lynch, 2016; Kappel, 2021; Lagewaard, 2021).

[17] See (Ranalli, 2020) for a helpful disambiguation of the term 'rationally irresolvable'. Consult (Martin, 2021c) for an argument *against* the rational irresolvability of deep disagreements. Finally, see (202X) for a recent case study in "conceptual engineering", where Guido Melchior suggests replacing discussions of deep disagreement with an analysis of rationally irresolvable disagreement because the latter notion can arguably be more clearly defined than the former while still capturing the basic intuitions underlying deep disagreement.

Table 1.2: Martin's Taxonomy of Deep Disagreements (Martin, 2021c, p. 5)

Breeds of Deep Disagreement	Complete	Partial
Strong	DD-StC	DD-StP
Weak	DD-WkC	DD-WkP
Distant	DD-DsC	DD-DsP

> ...explicit defenses of such principles will always be subject to a charge of circularity. Hume showed that the principle of induction is like this: you can't show that induction is reliable without employing induction. It also seems true of observation or sense perception. It seems difficult, to say the least, to prove that any of the senses are reliable without at some point employing one of the senses. Similarly with the basic principles of deductive logic: I can't prove basic logical principles without relying on them. In each case, I seem to have hit rock bottom... (Lynch, 2016, pp. 250-251)

As should be clear—in the case **Young Earth Creationist**—Henry and Richard disagree about the age of the Earth at surface-level, but their disagreement depends on a much more fundamental disagreement about evidential standards and what justifies beliefs. This is why their story has come to be viewed as a paradigmatic case of deep disagreement.

Surprisingly little has been written about the interconnections between logical and deep disagreement. Ben Martin's paper entitled '*Searching for Deep Disagreement in Logic: The Case of Dialetheism*' is an important exception. In this paper Martin observes that the domain of logic may be a fertile ground for deep disagreement:

> Much of our other knowledge requires us to presuppose that we possess certain logical knowledge or abilities. Consequently, it wouldn't be surprising if we were to find that there existed disagreements between competing schools of logic immune to rational resolution due to reaching the 'epistemic bedrock'. (Martin, 2021c, p. 2)

In order to carry out a case study on the depth of the disagreement between proponents of classical logic and dialetheism, Martin proposes his own **Taxonomy of Deep Disagreements** (captured by the table above). According to Martin, six different breeds of deep disagreement can be characterized when one combines the features specified in the table.

Consider first the leftmost column.

- ▷ *Strong* deep disagreement is a disagreement where S believes that $\langle p \rangle$, while S^* believes that not-$\langle p \rangle$, such that both S and S^* assume that their favored proposition is fundamental.
- ▷ *Weak* deep disagreement is a disagreement where S believes that $\langle p \rangle$, while S^* believes that not-$\langle p \rangle$, such that only one of S and S^* assumes that their favored proposition is fundamental.
- ▷ *Distant* deep disagreement is a disagreement where S believes that $\langle p \rangle$, while S^* believes that not-$\langle p \rangle$, such that neither S nor S^* assumes that their favored proposition is fundamental, but $\langle p \rangle$ is supported by a proposition S assumes to be fundamental and not-$\langle p \rangle$ is supported by a proposition S^* assumes to be fundamental.

Consider then the top row.

- ▷ *Complete* deep disagreement is a disagreement in which S and S^* don't agree regarding any of the propositions assumed fundamental by either party.
- ▷ *Partial* deep disagreement is a disagreement in which S and S^* disagree with respect to some of the propositions assumed fundamental by either party, but agree on other propositions assumed fundamental by either.

Combinations of the described features lead to the various breeds of deep disagreement categorized in Martin's matrix.

Let's consider the most extreme examples from Martin's taxonomy—viz., DD-DsP and DD-StC—for illustrative purposes, and see if we can make sense of them in the context of logical disagreement.

> **Logical DD-DsP**. Suppose S is a classical logician while S^* is an intuitionist. Say that S and S^* disagree over an instance of Double Negation Elimination as S believes that $\langle \neg\neg p \vDash p \rangle$ holds, whereas S^* doesn't. We can assume that while the classical logician works with an ontological presupposition of a realm of abstract objects independent of thinking agents, the intuitionist advocates for constructive methods only and takes logic to be about mental constructions. Thus, the intuitionist adheres to provability (or assertability) rather than truth in an objective sense. Their disagreement is allegedly deep, distant, and partial. Since the target-proposition under dispute is merely an instance of a general logical principle, the disagreement is distant. That is to say, $\langle \neg\neg p \vDash p \rangle$

> is not fundamental to S nor to S^*. However, the disagreement is still deep as $\langle \neg\neg p \vDash p \rangle$ is logically dependent on the excluded middle $\langle \vDash \varphi \vee \neg\varphi \rangle$ which is fundamental to both parties (the classical logician accepts the law of the excluded middle, while the intuitionist rejects it). Further, the disagreement is only partial since everything which is intuitionistically valid is also classically valid,[18] so in effect the logicians will in all likelihood agree with respect to other fundamental propositions.

> **Logical DD-StC**. S is a proponent of classical logic who firmly believes in the existence of logical laws, while S^* is a logical nihilist (understood as the view that there are no logical laws). Now, assume that S and S^* disagree over the status of L, where S believes that L is a logical law, while S^* disbelieves it. We can (perhaps) suppose that the contents of their beliefs are inconsistent with each other, it's just that the nihilist holds an error theory about logical laws, i.e., there are counterexamples to all the apparent ones, while the classical logician believes that there are some genuine logical laws and that L is one of them. Their disagreement is allegedly deep, strong, and complete. For every L, which is assumed to be a fundamental logical law by S, S^* will reject it outright.[19]

Although we should give Martin credit for having detected and articulated some important potential connections between logical and deep disagreements, there are some serious problems with his taxonomy. Just to mention one, by Martin's account it's not possible to be in deep disagreement if one side of the dispute merely suspends judgement about proposition $\langle p \rangle$ and its fundamental status. But clearly some of the most famous cases of deep disagreement are of exactly this sort. Take for instance:

> **Visual Perception**. Let 'p' refer to the sentence '*Visual perception is a reliable belief-forming process*'. Percy believes that $\langle p \rangle$, and assumes that $\langle p \rangle$ is fundamental. Skeptic suspends judgement as to whether $\langle p \rangle$, and is agnostic about the fundamentality of $\langle p \rangle$.

This is nothing but a classical skeptical scenario. Percy, on the one hand, takes himself to have a host of justified beliefs about the external world (or so we can assume) due to the fundamental reliability of the belief-forming process visual perception, i.e., Percy believes that $\langle p \rangle$ is true and assumes $\langle p \rangle$ to be fundamental, while Skeptic, on the other, simply introduces doubt about the epistemic merits of visual perception. Importantly, it's not the case that Skeptic believes not-$\langle p \rangle$, but rather that Skeptic withholds judgment about the matter and thereby induces skeptical doubt.

[18] See for example (Priest, 2008, Chapter 6) for a proof.

[19] 'Fundamental' for the nihilist might simply refer to one single level of fallacious logical laws.

Visual Perception is perhaps one of the most widely discussed examples of deep disagreement in the entire history of philosophy, and yet it isn't captured by Martin's account.

Chapter 2

Uniqueness and Logical Disagreement

This chapter discusses the *Uniqueness Thesis*, a core thesis in the epistemology of disagreement. After presenting uniqueness and clarifying relevant terms, a novel counterexample to the thesis will be introduced. This counterexample involves *logical disagreement*. Several objections to the counterexample are then considered, and it is argued that the best responses to the counterexample all undermine the initial motivation for uniqueness.

Keywords

The Uniqueness Thesis; Rational Uniqueness; Logical Disagreement; Logical Evidence; Propositional Justification; Epistemic Permissivism; Peer Disagreement

1 Introduction

The *Uniqueness Thesis* (henceforth denoted 'UT') concerns a relation between a body of evidence, a doxastic attitude, and a proposition. Jonathan Matheson, a proponent of the thesis, defines UT as follows:

> (UT) For any body of evidence E and proposition [p], E justifies at most one doxastic attitude toward [p] (Matheson, 2011, p. 360).

UT features frequently in the epistemology literature[1], especially in the debate concerning peer disagreement—if two epistemic peers[2] disagree about a proposition $\langle p \rangle$, is it then possible that they are both justified in their doxastic attitudes toward $\langle p \rangle$? If UT is true, the answer is negative.

Importantly, there are in fact several non-equivalent definitions of UT in the literature. Thomas Kelly, for example, favors a formulation of UT saying that there is *exactly one* justified doxastic attitude given a body of evidence (Kelly, 2010, p. 119), while Matheson prefers *at most one*, as we have just seen. Matheson notes that in most cases there will be exactly one justified doxastic attitude given a body of evidence, but in some situations, there may be no justified doxastic attitude toward $\langle p \rangle$ whatsoever. This can arguably happen when one is not able to, or when it is simply not possible to, comprehend the proposition at hand.[3] If one takes (possible) comprehension of $\langle p \rangle$ to be a necessary condition for the existence of a justified doxastic attitude toward $\langle p \rangle$, then it seems most reasonable to use Matheson's weaker definition of UT. Thus, this is what we will assume here.

Further, we will adopt Matheson's assumption that the term 'doxastic attitude' can only refer to the following three possibilities: *belief that* $\langle p \rangle$; *disbelief that* $\langle p \rangle$; and *suspension of judgement with respect to* $\langle p \rangle$, i.e., the possibility space of attitudes that one can take toward a proposition $\langle p \rangle$ is exhausted by these three attitudes.[4]

Now, UT puts a constraint on the total number of doxastic attitudes that a body of evidence can justify toward a proposition. According to UT any body of evidence E justifies at most one doxastic attitude toward $\langle p \rangle$. In other words, according to UT, there exists no body of evidence E such that E justifies both belief and disbelief toward $\langle p \rangle$. Similarly, of course, the thesis implies that there exists no E such that E justifies both a (dis)belief in $\langle p \rangle$ and suspension of judgement with respect to $\langle p \rangle$. In the paper titled '*The case for Rational Uniqueness*', Matheson makes two further clarifying remarks about UT:

> (UT) [...] makes no reference to individuals or times since (UT) claims (in part) that who possesses the body of evidence, as well as when it is possessed, makes no difference regarding which doxastic attitude is justified (if any) toward any

[1] See for example (Conee, 2010; Matheson, 2011; Rosa, 2012, 2016; Kelly, 2014; White, 2014, 2023; Kopec and Titelbaum, 2016; Ross, 2021; Kauss, 2023)

[2] Roughly put, two agents in disagreement are epistemic peers when neither side is epistemically superior with respect to the proposition at hand, i.e., when the two are similar enough in all relevant factors such as evidence, track record, time constraints etc.

[3] See (Feldman, 2006) for a motivation of this view.

[4] This assumption is common in the contemporary literature, see for example (Kelly, 2010; Matheson, 2011; Rosa, 2012; Titelbaum, 2015, 2019). Note that some have argued that the doxastic attitude of *disbelief that* $\langle p \rangle$ is non-equivalent to that of *believing the negation of* $\langle p \rangle$. See (Smart, 2021) for a recent argument. Unless otherwise stated we'll simply take disbelief that $\langle p \rangle$ and believing the negation of $\langle p \rangle$ as equivalent attitudes in what follows.

> particular proposition by that body of evidence (Matheson, 2011, p. 360).[5]
>
> (UT) concerns propositional justification, rather than doxastic justification. That is, the kind of justification relevant to (UT) is solely a relation between a body of evidence, a doxastic attitude, and a proposition. How individuals have come to have the doxastic attitudes they have toward the proposition in question will not be relevant to our discussion. Further, individuals can be propositionally justified in adopting attitudes toward propositions which they psychologically cannot adopt [...] Importantly, it is not a necessary condition for being justified in believing p that one be able to demonstrate that one is justified in believing (Matheson, 2011, pp. 360-361).

The first of these quotes states that according to UT a given body of evidence E justifies exactly the same doxastic attitude (if any) towards $\langle p \rangle$, no matter the subject that assesses E and at what time this is done. In the second quote, Matheson distinguishes between *propositional* and *doxastic* justification, where the former is a relation between a body of evidence, a doxastic attitude, and a proposition, the latter concerns *how* a given individual came to adopt a specific doxastic attitude towards a proposition, i.e., doxastic justification is concerned with one's reasons for actually adopting a certain attitude toward $\langle p \rangle$. Doxastic justification presumes that a given individual has a certain attitude toward $\langle p \rangle$, and the question is then whether or not this individual has sufficiently good (epistemic) reasons to be justified in having that attitude.[6] When it comes to propositional justification, on the other hand, it is irrelevant whether any individual is ever concerned with $\langle p \rangle$; the crux of propositional justification is that a justification-relation between a body of evidence, a doxastic attitude, and a proposition holds, not whether any individual realizes this. Understood in this way propositional justification refers to an external relation, and an individual can accordingly be propositionally justified in a doxastic attitude towards $\langle p \rangle$ even though this individual has not adopted the relevant attitude psychologically. And hence, it is not necessary for a subject to be able to demonstrate or defend this given attitude towards $\langle p \rangle$ in order for it to be propositionally justified. Matheson tells us that UT is a thesis concerning propositional justification rather than doxastic justification.

[5] Note that while Matheson's statement of UT doesn't make reference to individuals (i.e., cognizers or human agents) at all, some authors have presented versions of uniqueness that do. Consider for example Titelbaum and Kopec's tripartite distinction between propositional, attitudinal, and personal uniqueness (Titelbaum and Kopec, 2019, p. 206). *Propositional Uniqueness.* Given any body of evidence and proposition, the evidence all-things-considered justifies either the proposition, its negation, or neither. *Attitudinal Uniqueness.* Given any body of evidence and proposition, the evidence all-things considered justifies at most one of the following attitudes toward the proposition: belief, disbelief, or suspension. *Personal Uniqueness.* Given any body of evidence and proposition, there is at most one doxastic attitude that any agent with that total evidence is rationally permitted to take toward the proposition.

[6] For accounts of the epistemic *basing relation*, which is often taken to be relevant for doxastic justification, see for instance (McCain, 2014; Carter and Bondy, 2019; Korcz, 2021).

2 Clarifications

Before we move on to consider the announced counterexample to UT, let us pause to further specify what is meant by 'justification' and 'evidence' in the rest of the chapter. We will deliberately stay on a high level of generality in order not to exclude too many accounts of justification and evidence from the later discussions in sections 3 and 4.

When using the term 'justification,' this use is restricted to the epistemic domain, we are not concerned with any practical issues whatsoever. So, in other words, our concern is with the justification of doxastic attitudes towards propositions. This kind of justification is taken to be regulated by epistemic norms, i.e., truth-conducive norms, and as indicated in §1, we are concerned with *propositional* justification rather than doxastic justification.[7]

Our use of the term 'evidence' assumes that we can all agree that evidence can stem from many different sources like direct visual perception, testimony from individuals or media, scientific experiments etc. The only constraints we will force on our understanding of evidence from the outset are: (1) evidence must be propositional (and thus truth-apt); (2) any piece of evidence must be true; (3) any piece of evidence must (at least in principle) be accessible to human beings; and (4) evidence should be supportive of doxastic attitudes, where 'support' may be interpreted probabilistically, but does not have to be.

(2) is arguably the most controversial among these four constraints. For our purposes, however, there is a very good reason for including this factivity condition. To see this, suppose that one could have false pieces of evidence in one's (total) body of evidence E. Then, given the further assumption that false evidence can support anything, we could easily have a situation where a true bit of evidence e_1 from E entails $\langle p \rangle$ and thus supports the belief that $\langle p \rangle$, while a false bit of evidence e_2 from E entails not-$\langle p \rangle$ and thus supports disbelieving that $\langle p \rangle$, making E inconsistent and "explosive". This would in effect trivialize the debate about UT; on this account of evidence UT is obviously false.[8] Hence, we should either accept that evidence is factive or we should deny that false evidence can support anything. For the rest of the chapter we will take the first option.

3 The Argument from Logical Disagreement

Consider now the following case against UT:

[7] The literature on epistemic justification is vast, but prominent examples of theories of justification can be found in (Goldman, 1979, 1986; BonJour, 1985; Feldman and Conee, 1985; Alston, 1989; Sosa, 1991; Williamson, 2000; Conee and Feldman, 2004). Note also Littlejohn's tripartite division of epistemic justification which includes *personal* justification as well as doxastic and propositional (Littlejohn, 2012, p. 5). According to Littlejohn, doxastic justification is sufficient for personal justification, but not *vice versa*.

[8] Thanks to Franz Berto for pressing this point about false evidence.

> **Logical Disagreement**. Two logicians, S_1 and S_2, are walking into an empty auditorium where they find a deduction written on a blackboard. S_1 and S_2 are simultaneously looking at the board. As it happens, S_1 is a classical logician, while S_2 is an intuitionist. Now, by definition, the deduction consists in a finite number of steps, so all steps of the deduction except for the conclusion C will serve as a common body of evidence E, i.e., a set of propositions that are represented in a language that both logicians fully comprehend. The central question is then whether E entails C. Suppose that C on line n is the result of applying DNE (double negation elimination) to not-not-C on line $n-1$.[9] As S_1 accepts classical logic, she also accepts the inference from not-not-C to C, while S_2 given her intuitionist convictions denies DNE as a general rule of inference and thus denies that C needs to come out supported by E.

In this case we have a situation in which two agents possess exactly (!) the same evidence (the propositions represented by lines $n-1$ on the blackboard), but they are justified in diverging doxastic attitudes towards the relevant proposition in question, namely C. We see that E justifies S_1 in her belief that C, while E justifies (at least) suspension of judgement regarding C for S_2 (as C is not necessarily supported by E). Thus, the case is a clear counterexample to UT as the number of attitudes that E justifies exceeds one.

Of course, as the reader will have noticed by now, the case is concerned with a special type of evidence, i.e., evidence of the completely formal type that we find in pure logic and mathematics. This means that the counterexample is narrow in the sense that it does not indicate the existence of counterexamples to UT among other types of evidence.[10] However, this will

[9] Using standard notation that isn't meant to favor any logical tradition, DNE is an inference from $\Gamma \vdash \sim\sim \varphi$ to $\Gamma \vdash \varphi$, where 'Γ' denotes a set of sentences in a given language, '$\vdash$' denotes deducibility from left to right, '$\sim$' denotes a negation operator, and 'φ' picks out a single sentence of the language. Some readers may point out that it is underspecified in the case above whether S_1 and S_2 disagree over an *instance* or a *schema* of DNE. This is true, but it will not make a significant difference to the main argument of the chapter. The crux is that the logicians genuinely disagree. For more elaborate discussions of genuine logical disagreement the reader should consult (Hattiangadi, 2018; Hjortland, 2022; Hattiangadi and Andersen, 202X)

[10] However, some epistemologists have suggested that there are counterexamples to UT among other types of evidence. Consider, for example, a case where S_1 and S_2 discuss which football team will win the national league this season. Suppose that their discussion takes place the day before the final match day, and at this point of the season only two teams can win; either team A or team B (not both). Suppose further that the only evidence available to the subjects is a certain newspaper statistic, which shows the scores of the season so far. According to this statistic, team A is in front of team B by the smallest possible margin. Now, S_1 is convinced that team A will take the championship due to the statistical support for this (they are ahead at this point). However, S_2 suspends judgement about who will be the champions as team A leads with the smallest possible margin and it is still possible for team B to make it. In such a case the proponent of UT should say that at most one of the subjects' doxastic attitudes is justified, but one might argue that this is wrong. In such borderline cases it may seem that at least two out of three doxastic attitudes could be justified. If this is right, we have a counterexample to UT using another type of evidence, i.e., empirical data. Find similar borderline cases in (Kelly, 2014, pp. 299-300). For a recent discussion of (merely) statistical evidence and its role in epistemology, see (Silva, 2023).

be completely irrelevant as long as we regard UT as a *general* epistemic principle. If the case holds, we will have a counterexample sufficient for rejecting UT.

Finally, before taking on some pressing objections to the Argument from Logical Disagreement, one further clarifying comment is called for. Note that the logical disagreement described above isn't simply a case where S_1 and S_2 are talking past each other because of equivocation about the meaning of the expression 'not', as Quine (1986, p. 81) would have it. The reason why we can rule this out is a certain "technique for arguing that an apparent conflict is a real one" due to Williamson (1988).[11] In (1982) Harris established that in a system of natural deduction with two different operators for negation—classical ('$\neg$') and intuitionist ('$\rightharpoondown$'), respectively—the biconditional $\neg\varphi \leftrightarrow \rightharpoondown\varphi$ becomes provable, for any formula φ. From this basis Williamson's technique requires us to ask whether (i) there are rules of inference governing both $\neg$ and $\rightharpoondown$, and (ii) whether such rules could allow classical and intuitionist logicians (like S_1 and S_2) to characterize negation as the unique operator obeying those rules (up to logical equivalence).

As it turns out, the answer to (i) is positive: both $\neg$ and $\rightharpoondown$ obey Ex Falso Quodlibet ('*EFQ*') and the Introduction Rule for Negation, ('N_{Intro}'). Let φ, ψ be well-formed formulas. Then a monadic operator $\sim$ obeys EFQ, N_{Intro}, and N_{Elim}, just in case the following two schemas are valid:

$$\frac{\varphi \qquad \sim\varphi}{\psi}\ {\scriptstyle EFQ} \qquad\qquad\qquad \begin{array}{c} \overset{(n)}{\varphi} \\ \vdots \\ \dfrac{\bot}{\sim\varphi}\ {\scriptstyle (n)\ N_{Intro}} \end{array}$$

Here, numerals in brackets, i.e., (n), serve two distinct purposes: they mark discharged assumptions; and they indicate at which point in the derivation assumptions are discharged.

The answer to (ii) is also positive. EFQ and N_{Intro} are jointwise strong enough to define any monadic operator obeying them (up to logical equivalence). To see this, let $\sim_1$ and $\sim_2$ be any two monadic operators obeying EFQ and N_{Intro}. The following derivation establishes the deductive equivalence: $\vdash \sim_1 p \leftrightarrow \sim_2 p$.

[11] Note that our exhibition of Williamson's technique follows the order of presentation found in (Rossi, 2023). We follow Rossi's lead as his presentation of the material is very clear and detailed.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\overset{(1)}{p}\quad\overset{(2)}{\sim_1 p}}{p}\,{\scriptstyle EFQ}\quad\dfrac{\overset{(1)}{p}\quad\overset{(2)}{\sim_1 p}}{\sim_2 p}\,{\scriptstyle EFQ}}{\bot}\,{\scriptstyle EFQ}}{\sim_2 p}\,{\scriptstyle (1)\ N_{Intro}}}{\sim_1 p\rightarrow\sim_2 p}\,{\scriptstyle (2)\ \rightarrow_{Intro}}\qquad\dfrac{\dfrac{\dfrac{\dfrac{\overset{(3)}{p}\quad\overset{(4)}{\sim_2 p}}{p}\,{\scriptstyle EFQ}\quad\dfrac{\overset{(3)}{p}\quad\overset{(4)}{\sim_2 p}}{\sim_1 p}\,{\scriptstyle EFQ}}{\bot}\,{\scriptstyle EFQ}}{\sim_1 p}\,{\scriptstyle (3)\ N_{Intro}}}{\sim_2 p\rightarrow\sim_1 p}\,{\scriptstyle (4)\ \rightarrow_{Intro}}}{\sim_1 p\leftrightarrow\sim_2 p}\,{\scriptstyle \leftrightarrow I}$$

As the answers to both (i) and (ii) are positive, Williamson (1988, p. 111) proposes a proof-theoretic argument showing that the disagreement between classical and intuitionist logicians over DNE is a genuine one, and not merely a verbal dispute. Summa: If there is only one monadic operator—up to logical equivalence—obeying both EFQ and N_{Intro}, then this must rule out the possibility that the classical and intuitionist logicians are merely talking past each other when disagreeing about whether it obeys DNE. Either the intuitionist is right and the classicist wrong (or *vice versa*). In any case, there cannot be a single logic with two negation operators only one of which obeys DNE.

4 Objections and Responses

As the case presented above will be very hard to accept for many readers (for various reasons), the rest of the chapter aims to motivate the argument from logical disagreement. The strategy here is simple. While discussing various objections to **Logical Disagreement**, it will become clear that the UT-proponent can only avoid the counterexample by undermining the initial motivation behind UT, i.e., explaining away the counterexample to UT will lead to an indirect defeat of the thesis. In the following, five objections to **Logical Disagreement** will be scrutinized (§§4.1-4.5). The first two will simply be rejected, the third will be found underdeveloped, and while the remaining two can actually explain away the counterexample to UT, this can only be done by undermining the motivation behind the principle.

4.1 Evidence is Contingent

> **Objection 1**. Even though the evidence E present in **Logical Disagreement** satisfies our four rudimentary constraints on evidence (cf. §2) as E is propositional, factive, accessible, and supportive, E is still not a genuine body of evidence. For only contingent propositions can be evidence. Thus, UT is not even applicable in **Logical Disagreement**.

First of all, there is no principle reason why necessary propositions such as the ones found in pure mathematics and logic cannot be counted as evidence. Propositions of logic and mathematics can clearly serve the supportive role of evidence very well, i.e., such propositions speak in favor of certain hypotheses in the strongest possible way (by entailment). Hence, if any proposition is able to justify a belief, it seems that pure logical or mathematical propositions are ideal candidates. Habit may dictate, perhaps leading back to acceptance of Hume's Fork, that some of us cannot see the point in taking purely formal premises of deductive arguments as evidence, but without further qualification this is obviously not a good argument for accepting such an exclusion in philosophical or scientific work. Moreover, accepting **Objection 1** leads to absurd consequences when we hold other plausible epistemic principles to be true. Take for example Timothy Williamson's principle $E = K$, i.e., evidence equals knowledge (2000, Chapter 9). If we accept that our evidence is coextensive with our knowledge, and that **Objection 1** holds, it directly follows that we cannot have pure mathematical or logical knowledge. To deny that we can and do have such knowledge would not only be absurd, it would be intellectual suicide.

4.2 Communication Breakdown

> **Objection 2**. The case **Logical Disagreement** misrepresents the interaction between classical logicians and intuitionists. Where the classical logician works with a philosophical presupposition of a realm of mathematical objects independent of the thinking subject (objects that obey the laws of classical logic and can stand in set-theoretic relations), this is radically different from the intuitionists who advocate for constructive methods and take mathematics to be about mental constructions. As a result of this schism, the two logicians in the proposed case would run into an insurmountable communication breakdown, i.e., the DNE-inference acceptable to the classical logician would not even be understandable to the intuitionist—it would be nonsense. To quote Brouwer: "*Let us now consider the concept: 'denumerably infinite ordinal number.' From the fact that this concept has a clear and well-defined meaning for both formalist and intuitionist, the former infers the right to create the 'set of all denumerably infinite ordinal numbers,' the power of which he calls aleph-one, a right not recognized by the intuitionist.*" (Brouwer, 1975) Something similar to what Brouwer describes in the interaction between diverse logical traditions in this quote occurs in **Logical Disagreement** with respect to DNE, i.e., the intuitionist does simply not comprehend the final step of the deduction on the blackboard. Thus, suspension of judgement is not a justified doxastic attitude for the intuitionist in this case; the supposed logical connection between E and C is gibberish to her. Rather, **Logical Disagreement** represents the kind of case where there is no justified doxastic attitude for the intuitionist to have. Hence, UT would be saved (at least the *at most one doxastic attitude*-version of the thesis). The case allows only one justified attitude, namely the attitude of the classical logician.

This objection overstates the divide between the classical and intuitionist traditions. Comprehension of classical logic is often presupposed in discussions of non-classical logical systems, e.g., as a metatheory. Indeed, it is stipulated in **Logical Disagreement** that the deduction found on the blackboard is written in a language that both logicians fully comprehend. We do not need more than noticing and appreciating this very stipulation in order to slide off the objection.

Further, we can strengthen this reply by noticing that it is not the case that when there is logical disagreement, one party has automatically misunderstood (or lacks) some concept. The disagreement may just be the result of one side having false beliefs. So, in **Logical Disagreement**, it need not be the case that the intuitionist (supposing that she got it wrong) lacks some concept about how negation works, or has misunderstood or changed its meaning. Negation means whatever it means, also in the intuitionist's mouth, she just has false beliefs about that meaning.[12]

4.3 Logical Monism

Now, let us turn to the more challenging objections.

> **Objection 3**. The evidence E does in fact justify exactly one doxastic attitude in **Logical Disagreement**, it is just that we do not know which attitude it is. For we do not know which logical theory is the "correct" model of logical consequence, but surely there is only one correct logic in the end. Thus, UT survives the case even though the logical disagreement between the classical logician and intuitionist leaves us in the dark with respect to which doxastic attitude is justified by E.[13]

This objection begs the question against *logical pluralists* (e.g., Beall & Restall-style), i.e., the view that there is more than one true (or correct) logic.[14] According to logical pluralists, there is not always a single answer to the question whether a proposition $\langle p \rangle$ logically follows from a set of propositions (premises), in some cases there are more than one correct answer. A rough motivation for logical pluralism is that theories of classical logic, relevance logic, intuitionistic logic etc., all have a rightful place in formalizing and restraining logical inference as various important aspects of our pre-theoretic notion of logical consequence can be explicated by each of these approaches to logic.

[12] A similar point is made by Williamson in (2007, Chapter 4).

[13] See, e.g., (Griffiths and Paseau, 2022) for a recent defense of *logical monism*.

[14] In principle, the objection also begs the question against *logical nihilism*, which is the extreme view that there is no true (or correct) logic at all (Cotnoir, 2018; Russell, 2018a).

Clearly, begging the question against the pluralist in this way merely relocates the tension from an infight between UT-supporters and -deniers to a clash between logical monism and pluralism, so it seems like a dissatisfying option. Of course, some UT-supporters might be happy to say that logical pluralism is false, and thus they will have a way to save their principle, but this strategy should be supported by strong independent reasons. It will not be enough for the UT-supporter to accept logical monism because it seems like the default position amongst epistemologists. Hence, **Objection 3** is underdeveloped as it stands, and UT-supporters opting for this way out have further work to do.

Developing the back and forth between logical monists and pluralists any further here would take us beyond the scope of this chapter, but the reader can find some useful references in the footnote below.[15]

4.4 Splitting the Evidence

> **Objection 4**. As S_1 and S_2 belong to two opposing traditions in logic and thus do not accept the same rules of inference, it is actually not the case that they possess the same evidence in the situation described. Surely, considered just as a set of (formal) propositions, the evidence is the same for both subjects, but due to the subjects' diverse logical backgrounds the evidence splits in two. The case really presents both E and E^*, where the acceptable inference rules of classical logic are tacitly accepted to induce E and the rules of intuitionist logic are tacitly accepted to induce E^*. No set of (formal) propositions supports anything pre-theoretically. Choosing a logical theory is necessary to even generate *logical* evidence. Pre-theoretically, the question of which doxastic attitude is supported by a body of logical evidence is empty. Hence, **Logical Disagreement** is not a counterexample to UT since each body of evidence only justifies one doxastic attitude.

Prima facie, this objection seems to have something going for it. Indeed, it might save UT seen as a general epistemic principle since at most one doxastic attitude can be justified per body of evidence. However, at the same time it undermines the initial appeal of UT. For if we need to choose a logical theory in order to even generate logical evidence, we get a kind of relativism with respect to logical evidence. To illustrate, take an arbitrary set of (formal) propositions. This set does not constitute a unique body of logical evidence, as would be natural to suppose, instead it constitutes as many different bodies of logical evidence as there are acceptable logical theories.

[15] For more on logical pluralism in the Beall & Restall-style, see, e.g., (Beall and Restall, 2000, 2006). Other kinds of logical pluralism can be found in (Carnap, 2014; Shapiro, 2014). For an extensive overview, see (Russell, 2019b).

This moves our discussion away from evidence—as the central topic—to a discussion of acceptable theories instead, but no such discussion should be relevant to UT. UT should not be true only relative to preferred theory. For let us remind ourselves of how strong a thesis UT really is: it concerns all bodies of evidence, no matter what subject possesses it, and no matter the time and circumstances.

The crucial point is that UT is supposed to motivate a certain response to peer disagreement, i.e., at most one peer can be justified in her doxastic attitude toward the target-proposition in such disagreements. But if logical evidence is relativized to preferred logical theory, the scope of UT is reduced drastically. You can now only share logical evidence with those from your own theoretical equivalence class, and there can be as many of those classes as there are acceptable logical theories. This kind of relativism is clearly not desirable for a UT-proponent, and thus saving UT using **Objection 4** turns out to be a Pyrrhic victory.[16]

However, some might hesitate to admit that **Objection 4** leads to evidential relativism regarding logical evidence, for it may be objected that E and E^* don't have the same epistemic status. There could be good and purely epistemic reasons for favoring E over E^* (or *vice versa*) the reply goes. As noted above, E is the body of evidence induced by the tacit acceptance of classical logic, while E^* is the result of tacitly accepting intuitionist logic, but surely logicians do not just accept any old theory of logic, they have epistemic reasons for accepting whatever theory they favor. Thus, S_1's *total* evidence pool may very well include evidence for accepting DNE, law of the excluded middle etc., which the intuitionist lacks. Similarly, S_2's *total* evidence pool may well include evidence for denying DNE, law of the excluded middle

[16] Other epistemologists have suggested that one way in which uniqueness might fail is if there is a plurality of methods (in a broad sense) which one could rationally use to generate evidence. Accordingly, the counterexample **Logical Disagreement** presented here, and our discussion about logical evidence being relativized to acceptable logical theories, might be subsumed under a broader style of argument against uniqueness, namely that UT fails because evidence (of various types) is relative to acceptable methods. For further discussion of this general style of argument, see for instance (Goldman, 2010; Hales, 2014).

Note also how the issues surrounding logical evidence and uniqueness relate to some more established debates about permissible *epistemic standards* (Titelbaum and Kopec, 2019). Plenty of formal epistemologists claim that a body of evidence supports a hypothesis only relative to a rational reasoning method, and since there are multiple, extensionally non-equivalent, rational reasoning methods available, there is not always an unambiguous fact of the matter about whether some evidence supports a particular hypothesis. Subjective Bayesianism, for example, could deny UT by appealing to legitimate differences in epistemic standards. In general, Bayesians hold that any rational agent's credences at a given time can be obtained by conditionalizing their *hypothetical prior* ('Cr_h') on their total evidence at that time. For a total body of evidence E and a hypothesis H, the evidence supports the hypothesis exactly when $Cr_h(H \mid E) > Cr_h(H)$. Here, facts about evidential support are relative to the hypothetical prior of the relevant agent, and we can plausibly think of an agent's hypothetical prior as capturing their epistemic standards. Some Objective Bayesians claim that there is a unique rational hypothetical prior, so, in their case—while evidential support is relative to the hypothetical prior—there is still at most one rational hypothetical prior, and so UT is true. Yet some Subjective Bayesians claim that multiple hypothetical priors are rationally acceptable. Thus, for them, two rational agents could have different hypothetical priors, i.e., different epistemic standards, and end up in situations where the same body of evidence E supports a hypothesis H for one of them while it doesn't for the other.

etc., which the classical logician does not have in her possession. Further, S_1's reasons may be better than S_2's ditto (or *vice versa*).

Although this worry is legitimate, it will not save UT. First, it is underspecified in the literature whether UT is meant to apply to the *total* bodies of evidence in this sense, i.e., including pieces of evidence supporting one's methods used to generate evidence. There are hints about the importance of evidence for evidence-generating methods in the literature on *deep disagreement*,[17] but usually such evidence is taken as background information, and thus not as included in whatever body of evidence is under consideration in standard disagreement cases. Thus, it is not clear what UT-proponents would say about cases involving such *total* bodies of evidence. Further, one could easily rewrite **Logical Disagreement** stipulating that the two logicians were (known) epistemic peers. Then, insofar as evidential symmetry is necessary for peerhood, this would exclude any evidence from the case besides the common evidence. Of course, one could then say that if S_1 is a classical logician and S_2 an intuitionist, they cannot be epistemic peers, but in that case, we are back to square one; logical evidence becomes relativized to your own theoretical equivalence class and relativism looms.

4.5 Individualistic versus Social Epistemology

> **Objection 5**. UT is most plausibly defended as an *intra*-personal thesis, but **Logical Disagreement** is an inter-personal case.
>
> Thomas Kelly distinguishes between *intra-personal* and *inter-personal* versions of UT:

UT_{Intra} Given that my evidence is E, there is some doxastic attitude D that is the only fully rational doxastic attitude for me to take towards proposition p [...] (Kelly, 2014, p. 307).[18]

UT_{Inter} Given evidence E, there is some doxastic attitude D that is the only fully rational doxastic attitude for anyone to take towards proposition p [...].[19]

> Only UT_{Intra} holds as a general epistemic principle; not UT_{Inter}.

[17] For detailed discussions of deep disagreement, see (Lynch, 2010, 2016; Kappel, 2012, 2021; Ranalli, 2020, 2021; Ranalli and Lagewaard, 2022a,b; Barker, 2023).

[18] Note that even though Kelly uses the term 'rational' instead of 'justified' in the quote above, it will not make any substantial difference for our purposes.

[19] See footnote 18.

This objection saves UT as a general epistemic principle *intra-personally*, but as should be clear, it also completely undermines the core motivation for the thesis, which is social. Instead of relativizing evidence to acceptable theories or methods as in **Objection 4**, E is now relativized to subjects, and an even worse kind of relativism is unavoidable.

We should agree that UT_{Intra} is true. Take a perceptual case. If subject S clearly sees that there is a computer in front of her on the table and this visual perception constitutes her relevant evidence, then under normal circumstances there will be at most one justified doxastic attitude for her to adopt towards the proposition expressed by the sentence '*There is a computer on the table*', i.e., S is justified in believing the proposition to be true (while either disbelieving or suspending judgement would be unjustified). Likewise, UT_{Intra} seems true in logic cases insofar as we assume the agent in play has accepted a certain logical theory (as the only correct one) in advance. This blocks cases where **Logical Disagreement** is reformulated as a single person-case with an eclectic logician who prefers neither the classical nor intuitionist tradition of logic, and yet is fully competent in both. Given our assumption, this logician cannot be intra-personally justified in more than one doxastic attitude towards a given $\langle p \rangle$, e.g., the eclectic logician cannot be justified in a belief that $\langle p \rangle$ as well as a suspension of judgement with respect to $\langle p \rangle$ based on the same body of logical evidence.

However, as mentioned above, admitting that only UT_{Intra} is true comes with an unbearable cost for the UT-proponent. For with the embrace of this view, UT is no longer relevant to the peer disagreement debate which it was supposed to be central to. As UT_{Intra} is compatible with multiple doxastic attitudes being justified in cases of peer disagreement, the initial motivation behind UT is now completely lost. Thus, UT-proponents should not accept **Objection 5** as it indirectly undermines UT.

5 Concluding Remarks

This chapter has introduced a new counterexample to UT which involves logical disagreement. To legitimize this example and strengthen the case for it, we have shown that five different objections trying to save UT from **Logical Disagreement** fails. Two of the five objections were simply fended off, one needed further development to pose any real threat, while explaining away the counterexample with either one of the remaining two options resulted in an unbearable indirect defeat of the thesis. Hence, in the absence of successful objections to **Logical Disagreement**, the chapter recommends that we hesitate in accepting UT as a general epistemic principle.

Chapter 3

Second Preamble

1 Preamble

This second preamble aims to set the stage for Chapter 4. In §§1.1-1.4 the reader will find some useful information about the Theory Choice Reading of 'logical disagreement', Anti-Exceptionalism about Logic, Kripke's Adoption Problem, and Wittgensteinian "hinges".

1.1 'Logical Disagreement'—The Theory Choice Reading

As will be familiar to the reader at this stage of the monograph, the Theory Choice Reading of 'logical disagreement' is greatly inspired by the philosophy of science, where theory choice is an established topic. In a nutshell the interpretation says that genuine logical disagreements take place when entire logical theories come into conflict with each other; as opposed to disagreeing about sub-theoretic claims in a piecemeal fashion. On the Theory Choice Reading, logical disagreements concern how we justify our choice of a whole logical theory such as classical, intuitionistic, paraconsistent, connexive, quantum etc., rather than what we believe about a particular logical principle or inference, say, Modus Ponens.[1]

An illustrative case of logical disagreement—which is in line with Theory Choice Reading—concerns the dispute between classical mechanics and the quantum ditto. Ole Hjortland writes:

[1] To be sure, the term 'logical theory' must at minimum be understood as a set of sentences logically closed under a given entailment-relation, and according to Ole Hjortland there is something like a consensus that the main function of a logical theory is to tell us which inferences are valid (Hjortland, 2019, p. 252). However, some authors add to this deflationary understanding a demand that theories should account for features like provability, truth-preservation, formality, and consistency, as well (Priest, 2005; Hjortland, 2017).

> In classical mechanics, physical events are represented mathematically by subsets of phase spaces, i.e. coordinates for position and momentum. The physical events together with set-theoretic operations yield a Boolean algebra. If we think of the physical events as propositions, and, correspondingly, the set-theoretic operations as logical connectives, classical mechanics is governed by classical logic. In quantum mechanics, however, the mathematical representation of propositions as subspaces of a Hilbert space gives rise to a non-distributive lattice. With the meet, join, and orthocomplement operators interpreted as logical connectives, the result is a logic where the law of distributivity fails (i.e. $A \wedge (B \vee C) \nvDash (A \wedge B) \vee (A \wedge C)$)—more precisely, quantum logic. (Hjortland, 2019, p. 267)

In this context Hilary Putnam (1969) asserted that, if quantum mechanics were closed under classical logic, the result would be indefensible. From a cost-benefit analysis, he concluded that classical logic ought to be abandoned in favor of quantum logic, which he thought could avoid the indefensible results.[2]

As should be clear from the quantum mechanics-example, the Theory Choice Reading fits quite well with the idea that logic is continuous with (empirical) science. An idea which is sometimes referred to as 'Anti-Exceptionalism about Logic'.

1.2 Anti-Exceptionalism about Logic

There is a common perception of logic as an *exceptional* discipline in the sense of it being normatively, epistemically, methodologically, and metaphysically different from (empirical) science (Ferrari et al., 2023).

Ever since antiquity logical laws—like the Law of Identity stating that every entity is identical with itself—have been seen as special in many different ways, and logic has commonly been viewed as a foundational field of study underlying all other fields. Whereas the laws of physics apply only to physical systems, those of logic have typically been conceived as exceptionally general, i.e., in applying to all domains and entities. Logical laws have often been perceived as prescriptive across domains like geometry and physics, while the laws of geometry and physics haven't been viewed as prescriptive for logic. As such logic isn't concerned with the content of laws, sentences, principles, or propositions, but only with their *form*.

For this reason it has been the exceptionalist conception that principles of logic are necessary and analytic—i.e., not responsive to evidence from the empirical realm—as well as *a priori*, leading to traditional views like Rationalism and Semanticism (cf. Chapter 1, §1.3). According to such exceptionalist views of logic, it follows that logical evidence, justification, and

[2] For a critique of Putnam's revisionary argument in favor of quantum logic, see for example (Maudlin, 2005).

knowledge, must either stem from direct intuitions about the realm of logic or epistemic analyticity.[3]

Anti-Exceptionalism about Logic ('AEL') in its many forms challenges all the above mentioned exceptionalist aspects of logic. Historically, AEL has been associated with Quine (1951; 1986) who argued that logic is neither necessary, analytic nor *a priori*. Modern varieties of AEL, however, come in less radical forms, e.g., by denial of logic's a priori-status (Hjortland, 2017) and/or analyticity (Williamson, 2007) without full-blown Quinean commitments.

Much attention in recent debate has been paid to the question of whether logic is *epistemically* exceptional. Contrary to the traditional views—such as Rationalism and Semanticism—it has become increasingly popular to suggest that epistemic justification in the context of logic is a matter of showing that a given logical theory better accommodates the relevant data than its rivals (along with arguing for its possession of theoretical virtues and lack of vices). This view is now known as 'Logical Abductivism' and is summarized by Ben Martin as follows:

> According to this account of logical epistemology, logical propositions are not directly justified by intuitions or definitions, but rather logical theories are justified by their ability to best accommodate relevant data. In other words, logical theories are justified by abductive means. (Martin, 2021b, p. 9070)

Below—in Chapter 4—we'll present an argument against a necessary component of Logical Abductivism, viz.,*Justification Holism*.

According to Gillian Russell, *abductivists* endorse two central claims:

> The heart of the abductivist approach consists in two claims. The first is holism about the justification of logic: it is entire logics—rather than isolated claims of consequence—that are justified (or not). The second is that what justifies a theory is adequacy to the data, and the possession of virtues and absence of vices. (Russell, 2019a, p. 550)

For abductivists the object of justification is logical theories *en bloc* rather than individual claims of logical entailment. Abductivists endorse justification holism claiming that whatever justification we have for holding particular claims of logical entailment must be in virtue of the logical theory to which they belong. It's not that one is not able to have justification with respect to individual sentences about entailment, the point is rather that such justification is dependent on a choice of logical theory, say, classical, intuitionistic, paraconsistent, paracomplete etc.

[3] A sentence is *metaphysically* analytic if and only if it is true in virtue of meaning. We can contrast this with a sentence being *epistemically* analytic exactly when anyone who understands it is justified in taking it to be true.

Further Russell underscores that abductivism is incompatible with *Justification Atomism*:

> One view that is incompatible with abductivism is a view on which individual claims about entailment are justified atomistically, rather than in the context of a whole theory. (Russell, 2019a, p. 552)

The justification atomist opposes the holist part of the abductivist methodology by insisting that: *individual claims about entailment can be justified point-wise rather than in the context of a whole logical theory*.

Importantly, justification *holism* is not claiming that one cannot have justification for an individual claim that, say, 'double negation elimination is valid'.[4] For one could easily obtain such individual justification via a proof *within* some logical theory. The key point here is that, according to the holist, any such justification presupposes the context of an entire logical theory, and depends on a choice of such theory, e.g., choosing a classical theory rather than an intuitionistic one.

The *atomistic* view is incompatible with holism because the atomist holds that there can be cases of individual entailment-sentences such that these are justified outside the context of a whole logical theory, viz., counterexamples to holism.

As mentioned above we'll return to the theme of Logical Abductivism and its commitment to Justification Holism below in Chapter 4, where we'll argue that the holistic doctrine is false (cf. the Argument from Pre-Theoretic Universality).

1.3 Kripke's Adoption Problem

In addition to the themes of AEL, Abductivism, and Holism *versus* Atomism; Chapter 4 touches upon themes relating to Saul Kripke's so-called "Adoption Problem", which can be stated in dilemmatic terms in the following way:

> *Kripke's Adoption Problem*. Some basic logical principles cannot be adopted because, if a subject already infers with them, no adoption is needed, and if the subject does not infer in accordance with them, no (rational) adoption is possible.

The problem was first proposed in this form by Romina Padró (2015) and is thus also known as the 'Kripke-Padró Adoption Problem'.[5] Essentially the dilemmatic problem is a modern

[4] Let 'φ' denote a meta-variable and let the symbol '$\neg$' denote negation. Then double negation elimination is the entailment from $\neg\neg\varphi$ to φ.

[5] Note that Romina Padró has changed her name to 'Romina Birman' since then. We'll stick to the name 'Padró' here in order to avoid confusion.

reformulation of an argument, which was crafted by Kripke back in the 1970s, while he was engaging in a debate concerning logical theory choice and the very possibility of changing one's logic. Padró—in contrast—uses the modern version of the problem to induce pressure on a certain contemporary view in the epistemology of logic, viz., Inferential Cognitivism. This view claims that for any subject, S, logical inference is made possible by S's acceptance of basic logical principles. Padró's main idea is that if Inferential Cognitivism is correct, then we end up in the dilemma of the Adoption Problem. So, one way of avoiding the dilemma would simply be to reject the cognitivist stance.[6]

To illustrate the type of basic logical principles that is normally considered relevant to the Adoption Problem, let's remind ourselves of the famous allegorical dialogue between Achilles and the tortoise, as described by Lewis Carroll in (1895). Here, Carroll showed that there is an infinite regress problem concerning the inference rule Modus Ponens ('MP'). Using standard notation MP is stated thus:

$$\frac{(\varphi \to \psi) \qquad \varphi}{\psi}\ {\scriptstyle MP}$$

This rule tells us that a conditional together with its antecedent implies its consequent.

In Carroll's story (1895, p. 279) the problematic regress arises because Achilles is merely able to convince the tortoise that the two premises of MP are true in a concrete case (concerning two sides of a triangle being equal to each other), while he can't convince the tortoise to also make the inference from the truth of the premises to the truth of the conclusion of MP without appealing to a further conditional, viz., *if* $(\varphi \to \psi)$ is true and φ is true *then* ψ must be true. The crux of the problem is that unless the tortoise is willing to presuppose the "logical force" of MP, the tortoise can't apply the inference rule in concrete cases (like Achilles' example with the triangle). Achilles can only be the hero of the story insofar as the tortoise is willing to grant him that logic has a certain authority.[7]

[6] In this context it's important to distinguish between two different normative questions (as noted by Michael Devitt and Jillian Rose Roberts (202X)):

> *The Acceptance Question*. Can a subject rationally accept a new basic logical principle, accepting that inferences should be governed by a new rule R^* (perhaps replacing an old rule R)?
>
> *The Adoption Question*. Can a subject, on the basis of accepting a new basic logical principle, rationally change her practices so that her inferences are governed by R^* (perhaps instead of by an old rule R)?

Kripke's Adoption Problem is meant to provide a negative answer to the Adoption Question; not the Acceptance Question.

[7] See also (Quine, 2004) for a seminal discussion of Carroll's regress.

Below—in Chapter 4—we'll see that something similar might be the case with respect to the inference rule *Universal Instantiation* ('UI'), which essentially tells us that a universal quantifier implies any of its instances. Padró writes:

> Let's try to think of someone—and let's forget any questions about whether he can really understand the concept of "all" and so on—who somehow just doesn't see that from a universal statement each instance follows. But he is quite willing to accept my authority on these issues—at least, to try out or adopt or use provisionally any hypotheses that I give him. So I say to him, 'Consider the hypothesis that from each universal statement, each instance follows.' Now, previously to being told this, he believed it when I said that all ravens are black because I told him that too. But he was unable to infer that this raven, which is locked in a dark room, and he can't see it, is therefore black. And in fact, he doesn't see that that follows, or he doesn't see that that is actually true. So I say to him, 'Oh, you don't see that? Well, let me tell you, from every universal statement each instance follows.' He will say, 'Okay, yes. I believe you.' Now I say to him, '"All ravens are black" is a universal statement, and "This raven is black" is an instance. Yes?' 'Yes,' he agrees. So I say, 'Since all universal statements imply their instances, this particular universal statement, that all ravens are black, implies this particular instance.' He responds: 'Well, Hmm, I'm not entirely sure. I don't really think that I've got to accept that.' (2015, p. 49)

As in the example with Achilles and the tortoise above, it would seem that one needs to be able to reason with UI already in order to enable oneself to apply the rule in concrete cases (like the case concerning a specific raven in a dark room).

Notice, finally, how the Adoption Problem also has interesting connections to the debate concerning AEL and Logical Abductivism. That is to say, the problem might help us get a grip on the nature of logical inferences prior to any considerations about logical theory choice.

1.4 Hinges and Entitlement

For the remainder of the present preamble we shall be concerned with the topic of *Wittgensteinian hinge propositions*. A topic that relates to a number of themes in Chapter 4 in very puzzling ways.

In recent years *hinge epistemology* has become a hot topic in mainstream epistemology.[8] The term 'hinge epistemology' refers to a kind of epistemology where Wittgenstein's notion of

[8] See for example (McGinn, 1989; Moyal-Sharrock, 2004; Wright, 2004b,a; Fogelin, 2005; Coliva, 2010; Coliva and Moyal-Sharrock, 2016; Pritchard, 2010, 2021; Ranalli, 2020).

a *hinge* is central. This in itself is quite perplexing because it isn't really clear that such an epistemology is a genuine possibility to begin with.

The caption 'hinge' has roots in Wittgenstein's posthumous work *On Certainty* (1969b) and refers to a special kind of proposition, which is exempt from any doubt on the pain of outright disaster.[9] Hinges[10] are thus characterized by their maximal degree of resilience to counterevidence, although not as a result of critical thinking, empirical investigation, or philosophical inquiry (Wittgenstein, 1969b, §138). On the contrary, hinges are propositions that must be presupposed as a "scaffolding" of our critical thinking *tout court.* Various metaphors have been suggested in order to capture the special status of a hinge proposition, e.g., a scaffold, a riverbed, a ground, a pillar, a yardstick, a framework, a cornerstone etc. But whatever the metaphor, hinges should be understood as the content of some primitive certainties that we take for granted in our normal inquiries:

> The questions that we raise and our doubts depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which we turn. That is to say, it belongs to the logic of our scientific investigations that certain things are indeed not doubted... We just can't investigate everything, and for that reason we are forced to rest content with assumption. If I want the door to turn, the hinges must stay put. (Wittgenstein, 1969b, §§341-343)

Most insist on a disanalogy between primitive certainties and normal doxastic attitudes like beliefs, judgements, convictions, suppositions, seemings etc., and associate hinges with attitudes like presupposing, taking for granted, or holding fast, instead.

As Wittgenstein's original motivation for thinking about hinges was his attempt to solve the problem of radical skepticism about the external world, it is perhaps not an immense surprise that some widely used examples of hinge propositions are expressed by the sentences '*I am here*', '*Human beings have bodies*', and '*There are external objects*'. To cut a long story short, Wittgenstein's approach to the skeptical problem was to show that granting the certainty of some hinges was needed to even make sense of local doubting. Doubt on a global scale was senseless, according to him:

[9] Note that some prefer the term 'hinge commitment' to 'hinge proposition' as they don't regard hinges as genuine propositions that are capable of having truth value, see for instance (Moyal-Sharrock, 2004).

[10] There are at least four competing readings of *On Certainty* (and correspondingly at least four competing interpretations of hinges) in the literature (Coliva and Moyal-Sharrock, 2016). *The Therapeutic Reading*: Wittgenstein did not propose a substantive view of hinges, but merely aimed at curing our temptation of either doubting them or insisting on knowing them. *The Framework Reading*: Hinges are certain due to their crucial normative role in our lives (which completely exempts them from doubt). On this reading one could take on either a propositional or a non-propositional stance. *The Naturalist Reading*: Hinges are certain due to our *actual* practices. We take them for granted because of the communities we were actually brought up in. *The Epistemic Reading*: Hinges are special propositions that cannot be evidentially justified. However, on a broad conception of epistemic warrants that includes non-evidential ones, we can be *entitled* to accept them (for various reasons).

> If you tried to doubt everything you would not get as far as doubting anything. The game of doubting itself presupposes certainty. (Wittgenstein, 1969b, §115)

In line with this quote it's important to note two central interpretive claims that are often made about hinges. First, (i) hinges are "rule-like" entities (Wittgenstein, 1969b, §95; §98; §494) and they can be seen as propositions laying down certain "grammatical" rules that are needed to make sense of our language games (Coliva and Moyal-Sharrock, 2016, p. 15). The reason why we can't doubt them is not that we find it difficult, or even impossible, in the light of our evolution, psychological make-up, and socialization, *pace* the Naturalist Reading (cf. footnote 10). Rather we simply *cannot* have any reasons for doubting them—for reasons for doubting as such would presuppose the very hinges we're trying to doubt! Thus, second and relatedly, (ii) hinges are not evaluable by our normal epistemic standards, e.g., evidence, justification, knowledge etc. They are constitutive rules for the language games of believing and knowing rather than regular objects of belief and knowledge (Wittgenstein, 1969b, §4; §110; §§196-206).

For our purpose of discussing (the epistemic significance of) logical disagreement it is thus very interesting—and equally puzzling—to observe that also basic laws of logic have been seen as good candidates of being hinge propositions (Engel, 2016, p. 171). Since, given the interpretations of hinges we find in (i) and (ii), this would leave standard examples of logical disagreement as enigmas. How could we ever genuinely disagree about that which is impossible to doubt?

To exemplify the puzzle we have on our hands, consider again the quarrel over the Law of Non-Contradiction, i.e., the disagreement regarding the proposition $\langle\neg(\varphi \wedge \neg\varphi)\rangle$. Aristotle famously called this target-proposition the "firmest of all principles" (cf. the Metaphysics, Book IV). If the proposition expressed by $\neg(\varphi \wedge \neg\varphi)$ is indeed a hinge proposition, it wouldn't make much sense to (doxastically) disagree about it. After all—by (i) and (ii)—hinges are rule-like entities needed to even make sense of our language games, and thus they are rules for the language game of believing rather than feasible objects of belief.

Something like this might have been on David Lewis' mind when he wrote the following passage to Jc Beall and Graham Priest in reply to their invitation to contribute to a volume about the debate over the Law of Non-Contradiction:

> I'm sorry; I decline to contribute to your proposed book about the 'debate' over the law of non-contradiction. My feeling is that since this debate instantly reaches deadlock, there's really nothing much to say about it. To conduct a debate, one needs common ground; principles in dispute cannot of course fairly be used as common ground; and in this case, the principles not in dispute are so very much less certain than non-contradiction itself that it matters little whether

> or not a successful defence of non-contradiction could be based on them. (Lewis, 2004, p. 176)

Wright on Entitlement

As we said, hinge propositions are often considered ineligible to our normal epistemic standards. Prevailing epistemology is about how we justify beliefs, gain knowledge, revise doxastic attitudes in light of new evidence etc., but hinge propositions do not lend themselves to such issues—they seem isolated from them. Yet plenty of epistemologists have ironically enough tried to approach hinges from a characteristically epistemic angle. Amongst them we find Crispin Wright (2004b; 2004a), who has claimed that though we can't have regular epistemic justification *vis-à-vis* hinge propositions, we can still have a non-evidential kind of epistemic warrant with respect to them, viz., *epistemic entitlement* (see also (Burge and Peacocke, 1996; Dretske, 2000; Burge, 2003)).

Like Wittgenstein, Wright's philosophical work on hinges is motivated primarily by an attempt to fend off skeptical problems. His strategy is to argue that we have epistemic warrant for certain "cornerstone propositions," although that warrant doesn't have the normal form of evidential justification, the skeptic assumes it must have. According to Wright, we have entitlement to *accept*—rather than believe—certain propositions like those expressed by '*There is an external world*', '*There are other minds*' etc. based on *trust* alone. As such, epistemic entitlement to accept certain hinge propositions is a kind of rational warrant that doesn't depend on us having any evidence speaking for or against them. Rather we simply need to presuppose our entitlement to accept such hinge propositions since our normal justified beliefs concerning regular propositions are ultimately based upon those prerogatives.

In (2004b) Wright ends up with a view according to which we have entitlement to rationally trust certain foundational propositions, where *trust* is meant as a kind of acceptance that's weaker than belief, but stronger than mere acting on assumption. Roughly speaking there are two different types of epistemic entitlement by Wright's lights: (a) strategic entitlement; and (b) entitlement to a cognitive project. (a) occurs when subject S accepts a foundational proposition $\langle p \rangle$ in order to obtain a certain epistemic good; while (b) occurs when S accepts a foundational proposition $\langle p \rangle$ in order to carry out a given cognitive project.

While we shall not dwell on Wright's potential solution to skepticism here, it should be noted that some have found it very hard to see how any of the different versions of epistemic entitlement proposed by Wright are genuinely *epistemic* in nature. Carrie Jenkins (2007) has, for example, argued that even if we grant Wright that it is in fact rational for us to accept certain cornerstone propositions based on trust alone, this still can't be in virtue of proper epistemic rationality. Rather it must be by way of *instrumental* (or *practical*) rationality. To illustrate the alleged difference between instrumental rationality, on the one hand, and typical epistemic goods like rational belief, justified belief etc., on the other, consider the following case:

> John Doe is a brilliant set theor[ist] who is on the cusp of proving the Continuum Hypothesis: all he needs is six more months. But, alas, poor John is suffering from a serious illness that, according to his doctors, will almost certainly kill him in two months' time. John stubbornly clings to a belief that he will recover from his illness, and not only does this belief comfort him, but—let us suppose—it in fact significantly raises the chances that he will live for the six months that he needs both to complete his proof and to derive from it a variety of consequences for the rest of set theory. In other words, John's belief that he will recover is a causal means to his procuring a large number of true set-theoretic beliefs sometime in the future. But is John's belief epistemically justified? Is it the kind of belief that, from a purely epistemic perspective, he should be holding? (Berker, 2013, p. 369)

In response to this example it is usually claimed that John's belief is not epistemically justified because it isn't the sort of thing that John should believe for purely epistemic reasons. Rather John should believe in his own recovery due to the fortunate consequences that are likely to result from it. But this exemplifies a kind of instrumental means-end rationality rather than a genuinely epistemic kind (at least according to the standard view). By the golden standard of mainstream epistemology, epistemic rationality is not a matter of garnering a lot of epistemic value in the long run but rather a question of one's current epistemic backing with respect to some proposition of interest. A host of similar cases can be found in the literature, all featuring more or less fanciful imaginaries making it true that if an agent adopts a belief for which there is no good evidence (or other epistemic backing), then overwhelmingly good (epistemic) consequences will follow now and/or in the future.

In line with this one might think that it's somehow *instrumentally* rational for us to accept certain foundational propositions due to the downstream consequences, while still considering it doubtful that we could ever have genuinely *epistemic* entitlement to accept them.

Chapter 4

Countering Justification Holism in the Epistemology of Logic: The Argument from Pre-Theoretic Universality

A key question in the philosophy of logic is how we have epistemic justification for claims about logical entailment (assuming we have such justification at all). Justification holism asserts that claims of logical entailment can only be justified in the context of an entire logical theory, e.g., classical, intuitionistic, paraconsistent, paracomplete etc. According to holism, claims of logical entailment cannot be atomistically justified as isolated statements, independently of theory choice. At present there is a developing interest in—and endorsement of—justification holism due to the revival of an abductivist approach to the epistemology of logic. This chapter presents an argument against holism by establishing a foundational entailment-sentence of deduction which is justified independently of theory choice and outside the context of a whole logical theory.

Keywords

Deduction; Semantics; Bootstrapping; Logical Theories; Epistemic Justification; Logical Abductivism; Anti-Exceptionalism about Logic

1 Introduction

1.1 Abductivism, Justification Holism, and Logical Theories

Recently there has been a renewed interest in an abductivist approach (to be defined) in the epistemology of logic.[1] Some of the contemporary abductivists are motivated by *anti-exceptionalism about logic*, which, roughly speaking, says that logic doesn't differ from (empirical) science in any interesting way.[2] This view, and the general abductive approach, has historically been associated with Quine (1951, 1986),[3] [4] who argued that logic is neither necessary, analytic nor *a priori*.[5] Modern varieties of anti-exceptionalism, however, come in less radical forms, e.g., by denial of logic's a priori-status (Hjortland, 2017) and/or analyticity (Williamson, 2007) without full-blown Quinean commitments.

According to Gillian Russell, *abductivists* endorse two central claims:

> The heart of the abductivist approach consists in two claims. The first is holism about the justification of logic: it is entire logics—rather than isolated claims of consequence—that are justified (or not). The second is that what justifies a theory is adequacy to the data, and the possession of virtues and absence of vices. (Russell, 2019a, p. 550)

For abductivists the object of justification is logical theories *en bloc* rather than individual claims of logical entailment.[6] [7] Abductivists endorse justification holism claiming that what-

[1] See (Priest, 2005, 2014, 2021; Williamson, 2007, 2017b, 2020a, 202X; Russell, 2014, 2015, 2019a; Beall, 2017, 2019; Hjortland, 2017, 2019, 2022; Martin, 2021c,a,b, 2022; Zanetti, 2021; Martin and Hjortland, 2021, 2022; Rossberg and Shapiro, 2021; Sagi, 2021; Becker Arenhart, 2022a,b; Carlson, 2022; Tajer, 2022b; Ferrari et al., 2023; Martin and Hjortland, 202X).

[2] In a recent paper Martin and Hjortland (2022) distinguish between different kinds of anti-exceptionalism about logic. Usually anti-exceptionalism is taken to be a stronger claim than abductivism, e.g., *methodological* anti-exceptionalism proposes a similarity between the methodology in logic and science which is not necessary for abductivism.

[3] One should not simply identify modern versions of abductivism with Quine's ditto. See for instance (Martin, 2021b) for some important differences.

[4] Note also the seminal work on the abductive approach by Nelson Goodman (1983).

[5] Bear in mind the internal tension in (the development of) Quine's philosophy. On the one hand, *Quine the holist* (1953) takes logic to be revisable, it's just that our beliefs concerning such matters are closer to the center of our web of beliefs, and hence hard to revise, whereas beliefs about "more synthetic" statements are closer to the periphery of the web, and thus easier to revise. On the other hand, *Quine the conservative* (1986) thinks that classical first-order logic is "the realm of the obvious" and that any attempt of non-classical revision amounts to changing the subject.

[6] We'll use the terms 'proposition', 'sentence', and 'claim' interchangeably throughout this chapter.

[7] It's unclear in the contemporary literature on abductivism whether we should distinguish between a *logic* and a *logical theory*. Consult (Mortensen, 2013) for an example of someone who draws a clear distinction between the two.

ever justification we have for holding particular claims of logical entailment must be in virtue of the logical theory to which they belong. It's not that one is not able to have justification with respect to individual sentences about entailment, the point is rather that such justification is dependent on a choice of logical theory, say, classical, intuitionistic, paraconsistent, paracomplete etc.[8] Further, abductivists hold that the grounds for justification of a logical theory is how well it fits with relevant data (frequently taken to be our intuitive judgments about logical inferences) plus its theoretical virtues and lack of vices, e.g., its strength in terms of ratified consequences (in logic and wider scientific context), how aesthetically elegant and simple it is, and how ontologically parsimonious.

Abductivism is succinctly summarized by Ben Martin:

> According to this account of logical epistemology, logical propositions are not directly justified by intuitions or definitions, but rather logical theories are justified by their ability to best accommodate relevant data. In other words, logical theories are justified by abductive means. (Martin, 2021b, p. 9070)

To be sure, the term 'logical theory' must at minimum be understood as a set of sentences logically closed under a given entailment-relation (modeling the concept of validity). Indeed, according to Ole Hjortland there is something like a consensus that the main function of a logical theory is to tell us which inferences are valid (Hjortland, 2019, p. 252). However, some authors add to this deflationary understanding a demand that theories should account for features like provability, truth-preservation, formality, and consistency, as well (Priest, 2005; Hjortland, 2017).[9]

One should also bear in mind that, in some cases, e.g., Carnap (2014), Dummett (1991), and

[8] Further details about the position *justification holism* can be found in §2.3 below.

[9] As an anonymous reviewer points out, the minimal characterization of a logical theory stated above can be thought to miss a potential distinction between a logical *system* and a logical *theory*; where the former is taken to be a formal apparatus with a vocabulary, a proof-theory, a semantics etc., while the latter is an applied system that models particular target-phenomena. According to some, logical theories should not only tell us which inferences are valid, but ideally also tell us *why* these inferences are valid (and other inferences invalid). Theories shouldn't merely give us a set of sentences logically closed under a given entailment relation or supply a list of inferences or laws that are valid, they should also provide an account of why these inferences or laws are valid. Thus—according to some—logical theories are *about* a particular (extra-systematic) subject matter, and for a theory to be *correct* it should get the subject matter right. For example, the modal logics **S5**, **S4**, etc., can be characterized as logical systems of sets of sentences, given by some system of proofs or models. But to adopt one of these as a *theory* is to, in addition, adopt this or that system as part of an explanation of what follows from what (and what doesn't follow). You wouldn't adopt both **S4** and **S5** as logical theories of the same phenomenon (say, some given notion of necessity), when they give different accounts of the truth of modal statements that can differ in truth value. Note that while this distinction between logical system and logical theory is a plausible one, it won't change the main result of the present chapter whether we commit to it or not. See for instance the discussion of free logic below for an explanation how the main argument of §2 is compatible with different logical analyses.

Shapiro (2014), logical theories are claimed to be solely about language, i.e., metalinguistic, but often they are taken to be non-metalinguistic (Russell, 2009; Sider, 2013; Maddy, 2014; Williamson, 2013c, 2017b). Tim Williamson, for instance, takes logical theories to consist of unrestricted generalizations about the world, not just language.[10]

1.2 Justification Atomism

For the present purposes it's crucial to note that abductivism is incompatible with *justification atomism*:

> One view that is incompatible with abductivism is a view on which individual claims about entailment are justified atomistically, rather than in the context of a whole theory. (Russell, 2019a, p. 552)

The justification atomist opposes the holist part of the abductivist methodology by insisting that: *individual claims about entailment can be justified point-wise rather than in the context of a whole logical theory*.

Importantly, justification *holism* is not claiming that one cannot have justification for an individual claim that, say, 'double negation elimination is valid'.[11] For one could easily obtain such individual justification via a proof *within* some logical theory. The key point here is that, according to the holist, any such justification presupposes the context of an entire logical theory, and depends on a choice of such theory, e.g., choosing a classical theory rather than an intuitionistic one.

The *atomistic* view is incompatible with holism because the atomist holds that there can be cases of individual entailment-sentences such that these are justified outside the context of a whole logical theory, viz., counterexamples to holism.

Of course, some holists may be more sensitive to counterexamples than others. Tim Williamson's work on the problem of overfitting in epistemology (2007; 2017a; 2020a) suggests that he would be reluctant to give up holism due to a single counterexample, for instance; while Gillian Russell's work on logical nihilism (2017; 2018a; 2018b) indicates that she has a great respect for the normative force of individual counterexamples. Accordingly, the announced

[10] One might frame anti-exceptionalism about the content or subject matter of logical theories as *metaphysical* anti-exceptionalism about logic. An illustrative example of such position is Bertrand Russell's universalism: "*Logic is concerned with the real world just as truly as zoology, though with its more abstract and general features.*" (Russell, 1919, p. 169). Metaphysical anti-exceptionalism is importantly distinct from *epistemological* anti-exceptionalism, and as noted by Martin and Hjortland (2022), one can be an anti-exceptionalist about one without being an anti-exceptionalist about the other.

[11] Let 'φ' denote a meta-variable and let the symbol '$\neg$' denote negation. Then double negation elimination is the entailment from $\neg\neg\varphi$ to φ.

argument against justification holism (cf. §2) will have the greatest impact on those who are ill-disposed to counterexamples.

It's also worth stressing that the contemporary abductivists are not always explicit about what kind of epistemic justification they are interested in, and whether this is a kind that only logical experts can possess. *Prima facie*, the kind of justification one can expect agents to have with respect to logical propositions and theories varies with their logical background knowledge. Contrast, for example, the kinds of justification we would expect a novice and a logical expert to have, respectively. The expert may have firm convictions regarding logical theories and principles, while it's unlikely that the novice would even fathom what a logical theory is. However, it seems that the distinction between *fundamental* and *non-fundamental* sources of justification could dissolve this issue. Deductive proofs may be seen as a fundamental source of justification, while testimony could be considered a non-fundamental source enabling transmission of justification only. Insofar as we are interested in fundamental justification alone, it is straightforward to suppose that the justification of entailment-sentences is an esoteric business of logical experts, and that is what we will assume here.

Further, we'll suppose that the abductivists are interested in *propositional* rather than *doxastic* justification, i.e., the justification of logical propositions rather than belief-tokens about such propositions. Doxastic justification is a property that a belief has when one believes a proposition for which one has propositional justification, and this belief is based on that which propositionally justifies it. We will focus on propositional justification since—assuming we can give a good account of propositional justification and that this account can be exploited as the basis for the relevant beliefs—we can have doxastic justification as well.[12]

1.3 E-Sentences and E-Literals

Before getting down to business it will be helpful to introduce some technical terminology concerning logical entailment. *E-sentences* are atomic sentences in which the main predicate is given by the symbol '$\models$' (or its natural language equivalents) (Russell, 2019a).[13] Examples

[12] We'll leave it as an open question whether the distinction between justification *internalism* and *externalism* is of great importance to the holist. Note, however, that basing your beliefs about logical propositions on proofs in deductive logic could be seen as a kind of (evidential) proper basing of propositional justification, which would amount to doxastic justification on standard internalist accounts. Similarly, forming your beliefs about logical propositions via proofs in deductive logic could be counted as a reliable (or safe) method of belief-formation on standard externalist accounts of doxastic justification. For details on internalism in the form of evidentialism, consult, e.g., (Feldman and Conee, 1985; Conee and Feldman, 2004). For accounts of the epistemic basing relation, see, e.g., (McCain, 2012, 2014; Carter and Bondy, 2019; Neta, 2019; Korcz, 2021). For details regarding externalism in the form of process reliabilism, consult, e.g., (Goldman, 1979, 1986). For externalist accounts involving modal properties like safety and sensitivity, see, e.g., (Dretske, 1971; Nozick, 1983; Williamson, 2000; Pritchard, 2005).

[13] 'E-sentence' is shorthand for 'entailment-sentence'. As indicated by (the standard use of) the double turnstile-symbol '$\models$', E-sentences and E-literals should be thought of in semantic terms, not proof-theoretic ones

are:

- $[\varphi \vee \psi, \neg\psi \vDash \varphi]$
- $[\vDash \neg(\varphi \wedge \neg\varphi)]$
- $[\varphi \wedge \neg\varphi \vDash \psi]$[14]

These sentences are *atomic* in the sense that they are the simplest kind of sentences of a given meta-language. To see this, we observe that symbols like '$\vee$','$\neg$', '$\wedge$' are not used but merely mentioned in E-sentences, whereas '$\vDash$' is a metalinguistic symbol placed between terms referring to schemas (or sentences) of an object-language.[15]

An *E-literal* is either an E-sentence or its negation. Thus, all E-sentences are E-literals, but not *vice versa*. Examples of E-literals are:

- $[\varphi \rightarrow \psi, \varphi \nvDash \psi]$
- $[\vDash \varphi]$
- $[\varphi \wedge \neg\varphi \nvDash \psi]$

E-literals are central to the epistemology of logic as their truth-value tells us what follows from what, and what doesn't follow. On the common view that logic is the study of (valid) inferences, the importance of E-literals is given, but in virtue of what are our E-literals justified, and is it possible for individual E-literals to be propositionally justified outside the context of a whole logical theory? Those are the central questions of this chapter.[16] Justification holism gives one possible all-encompassing answer, but as we shall see now, there are good reasons to think that holism is false.

(more on our exclusive semantic focus in footnote 20). We use square brackets around entire E-sentences and E-literals rather than corner-quotes around schemas to ease readability.

[14] Let lowercase Greek letters be meta-variables. Let the symbols '$\vee$','$\neg$', '$\wedge$', and '$\rightarrow$', denote disjunction, negation, conjunction, and material implication, respectively.

[15] Note that our use of the object-language/meta-language distinction presupposes that there is a hierarchy of languages in logic. A number of logicians reject this. Notoriously, they think (i) it's implausible that there be meta-languages for English or any other natural language, and (ii) one does not even need a hierarchy of languages for the purposes of a theory of truth. Examples are dialetheists, such as Graham Priest (2006) and Jc Beall (2011), as well as proponents of paracomplete logics like Saul Kripke (1976). These logicians endorse non-hierarchical truth theories and semantics. It's well beyond the scope of this chapter to go deeper into these issues, so we'll have to make do with the following observation. Look in any logic textbook and you shall find a formal object-language plus a logical entailment-relation for that object-language defined in a meta-language, which is usually English (perhaps with bits of mathematical notation). In this sense, the notion of logical entailment is clearly meta-linguistic.

[16] Note that this is a separate question from the question of what makes an agent entitled in her *disposition* to reason in accordance with some rule (Boghossian and Peacocke, 2000).

2 A Foundational E-Sentence of Deduction

This section aims to show that the E-literal $[\forall x Px, \Gamma \vDash Pa]$, where '$a$' refers to an element of domain D of some model $\mathfrak{M}$, and 'Γ' denotes a (possibly empty) set of side-conditions, is true under any acceptable deductive entailment-relation, and denying its truth would mean giving up on deduction altogether.[17] In other words, the aim is to establish that a liberal version of the E-literal about *universal instantiation* is a foundational E-literal for which we have propositional justification independently of theory choice and outside the context of an entire logical theory; thus constituting a counterexample to the holistic doctrine.[18]

The plan for the rest of the section is as follows. In §2.1 universal instantiation is defined and some crucial notions, viz., *Universality* and *Universality Booting*, are introduced and motivated.[19] In §2.2 the main argument against justification holism is put forward. If successful, it shows that justification holism is false. As this result will strike many readers as being too bold, §2.3 aims to address some objections to it. In particular, the straightforward objection from *free logic* will be discussed in §2.3.

2.1 Terminology and Lemmas

Some preliminary remarks.

First, universal instantiation ('UI') is a well-known syntactic inference rule. Under one plausible semantic interpretation it says: any instance of '*Everything is P*' entails '*t is P*', where 't' refers to an individual term. When the rule is stated formally in standard notation, it looks like this:

$$\frac{\forall v P v}{P t}$$

When this schema is interpreted in the standard way, we take the quantifier denoted by '$\forall$' as ranging over a domain of objects, the predicate denoted by 'P' as referring to a property, and the term denoted by 't' as replacing all occurrences of the variable given by 'v'. Accordingly, we can state an E-literal about UI as follows: '$[\forall x Px, \Gamma \vDash Pa]$', where '$a$' refers to an ele-

[17] A (Tarskian) model $\mathfrak{M}$ in first-order logic is an ordered pair $\mathfrak{M} = \langle D, I \rangle$, where D is a domain of objects and I is an interpretation function specifying referents for constant symbols, predicate symbols, and function symbols. We say that $\mathfrak{M}$ is a model of a well-formed formula φ if φ is true in $\mathfrak{M}$. A countermodel $\mathfrak{M}^*$ to φ is a model of $\neg\varphi$.

[18] From this point on we'll frequently use the adjective 'foundational' about a particular E-literal and simply take this to mean *an entailment claim for which we have propositional justification independently of theory choice and outside the context of an entire logical theory*.

[19] 'Universality Booting' is shorthand for 'Universality Bootstrapping'.

ment of domain D of some model $\mathfrak{M}$, and 'Γ' denotes a set of side-conditions based on one's favored logical analysis. Since Γ is usually left empty, we'll simply write '$[\forall x Px \vDash Pa]$' by default in order to ease readability. We'll discuss a special case where Γ is non-empty in §2.3.

Second, we'll assume that, in semantics, *Universality* is a necessary property of every acceptable deductive entailment-relation. That is to say, any acceptable deductive entailment-relation—modeling the concept of validity—must involve universal quantification over cases, be it in the form of possible worlds, constructions, situations, truth-makers etc. One could, for instance, say:

> A valid inference is one whose conclusion is true in every case in which all its premises are true. (Jeffrey and Burgess, 2006, p. 1)

Or

> ... [D]eductive validity can be adequately accounted for by means of quantification over possible worlds: an argument is deductively valid (or equivalently, the relation of consequence holds between its premises and conclusion) if and only if in all possible worlds in which the premises are true/holds, so is/does the conclusion. (Dutilh Novaes, 2020, pp. 14-15)

In these and similar ways universal quantification is standardly thought to be embedded in the semantic characterization of deductive entailment. And furthermore, Universality is widely thought to be exactly what gives deduction necessary force, i.e., demarcating it from induction and abduction (Beall and Restall, 2000, 2006; Cohnitz and Estrada-González, 2019; Dutilh Novaes, 2020; Douven, 2021). Thus, Universality is an extremely well-motivated property of acceptable deductive entailment.[20]

Third, let's make the crucial observation that the E-literal about universal instantiation, i.e., $[\forall x Px \vDash Pa]$, is a universal sentence about *true* universal sentences. For the main predicate of $[\forall x Px \vDash Pa]$ is given by the entailment-symbol, which is exactly a universal claim (by Universality). This is crucial because, in our modelings of the concept validity, we'll have that any model $\mathfrak{M}$ which makes $[\forall x Px \vDash Pa]$ true must itself be a fact of universal quantification over cases; and note that this fact will need to be a *pre-theoretic* counterpart of UI. That is to say, any $\mathfrak{M}$ making the E-literal $[\forall x Px \vDash Pa]$ true must itself be a fact of universal quan-

[20] A natural constraint on the main result below is imposed by our exclusive focus on semantic accounts of deduction. Proof-theorists need not adhere to universal quantification over cases in their modelings of validity as their definitions of the concept presuppose the particular *there's a proof* rather than the universal *in all cases*. Structurally, however, a similar foundational point could be made with respect to the particular quantifier, but we'll leave proof-theoretic specifications of validity out of the picture here, as they are strictly speaking irrelevant to the aim of this chapter.

tification which lies outside the bounds of logical theorizing; since any acceptable deductive entailment-relation—modeling the concept of validity—must adhere to brute universal quantification over cases. Or, in yet other words, the E-literal about UI is *doubly* universal in containing both a universal statement and in stating a fact of entailment, which is itself a brute fact of universal quantification.[21] Let's name this special feature of $[\forall xPx \vDash Pa]$ '*Universality Booting*'.

Here's an intuitive elaboration. Consider the following E-literals:

1. $[\forall xPx \vDash Pa]$
2. $[\varphi \wedge \psi \vDash \varphi]$

Now, (1) induces Universality Booting, whereas (2) doesn't bring about anything like "Conjunctive Booting". For (1) is a universal sentence about true universal sentences, while (2) is a universal sentence about true conjunction-sentences. Hence, while any $\mathfrak{M}$ making (1) true must itself be a pre-theoretic fact of universal quantification over cases, it would be false to suggest that any $\mathfrak{M}$ making (2) true must itself be a pre-theoretic fact of conjunction elimination. And consequently, the E-literal $[\forall xPx \vDash Pa]$ has a pre-theoretic booting-property which other E-literals like $[\varphi \wedge \psi \vDash \varphi]$, $[\varphi \vDash \varphi \vee \psi]$, $[\neg\neg\varphi \vDash \varphi]$ etc. don't have.

In slogan-form: *Whatever logical theory you prefer, it will be booting in a state of universality!*

2.2 Countering Justification Holism

Based on the preliminaries from §2.1, we are now equipped to show that $[\forall xPx \vDash Pa]$ is a foundational E-literal of deduction.

> *The Argument from Pre-Theoretic Universality*
>
> Assume that Universality is a necessary property of any acceptable deductive entailment-relation, and let '$\vDash$' denote any such relation. Suppose further that $[\forall xPx \vDash Pa]$ is false. Then there exists a counter-model $\mathfrak{M}^*$ to the E-literal $[\forall xPx \vDash Pa]$, i.e., a model such that $[\forall xPx \nvDash Pa]$ and $a \in D$. By Universality Booting, any $\mathfrak{M}$ making $[\forall xPx \vDash Pa]$ true is itself a pre-theoretic fact of universal quantification over cases. Yet, by assumption $[\forall xPx \vDash Pa]$ is false, so there can be no such pre-theoretic fact. But then, by Universality, $\vDash$

[21] The brute fact of universal quantification referred to above is perhaps easiest to register when thinking in terms of counterexamples. If you've got a model of the premises of an argument which is not a model of the conclusion. Then you are making a transition from an *instance* to the falsity of a *universal* claim. This is an *implicit* appeal to UI. Thanks to an anonymous reviewer for their very detailed comments on this section.

> cannot be an acceptable deductive entailment-relation. For there exists a counterexample to universal quantification over cases, viz., $\mathfrak{M}^*$. Therefore, either $[\forall x Px \vDash Pa]$ has no counter-model, or Universality is not a necessary property of acceptable deductive entailment. By assumption, Universality is a necessary property of acceptable deductive entailment. Ergo: $[\forall x Px \vDash Pa]$ is true under any acceptable deductive entailment-relation.

Cut your theoretical cake anyway you please, some E-literals—like $[\forall x Px \vDash Pa]$ as demonstrated—are propositionally justified independently of theory choice and outside the context of an entire logical theory. And importantly, the upshot is not just that all acceptable logical theories should include $[\forall x Px \vDash Pa]$, perhaps for different reasons, rather the argument shows that $[\forall x Px \vDash Pa]$ is foundational in such a way that it leaves any theoretical specifications—within the bounds of deduction—redundant with respect to its justificational status. If one were to deny the truth of $[\forall x Px \vDash Pa]$, this would amount to giving up on deduction altogether (by denial of Universality). So, to carve out the point: $[\forall x Px \vDash Pa]$ is a foundational E-literal of deductive entailment, and hence justification holism must strictly speaking be false.[22]

Now, finally, before taking on some pressing objections to the Argument from Pre-Theoretic Universality, two quick clarifying comments are called for.

First, the argument above doesn't fall prey to a conflation of the distinction between quantification in *object-language* and quantification in *meta-language*. The argument appeals to the brute fact that any acceptable deductive entailment-relation—semantically understood—will be booting up in a state of universality with respect to its cases, be it in the form of possible worlds, constructions, situations, truth-makers etc. As this fact must be taken for granted by any logical theory, it will need to be presupposed in whatever semantic entailment-relation one can come up with, and no matter the meta-language one might fancy.

Second, neither does the argument conflate *first-order* and *higher-order* quantification. It uses no quantification over properties at all (or anything in that vicinity).

[22] It's worth flagging that the argument relies on inferential strategies such as *reductio ad absurdum* ('reductio'), which is unacceptable to some non-classical logicians, e.g., dialetheists like Graham Priest (2006). However, even for dialetheists who reject reductio as a general strategy, it's still safe to use it in consistent contexts. For Priest, reductio is "quasi-valid", i.e., valid if the premises are consistent. So, while reductio is used in the argument above, it's fair to suppose that this is in a consistent context, and thus, that even Priest would be fine with this particular use.

2.3 Objections

Charity to Holists

One potential worry about the above argument concerns how one should interpret the position referred to by the label 'justification holism' and whether the result in §2.2 really poses a problem for the holist under a charitable interpretation.[23] In this chapter, the holist position was introduced as follows:

> (a) Holism about the justification of logic: it is entire logics—rather than isolated claims of consequence—that are justified (or not). (cf. §1.1)

But when countering this claim, it was established that:

> (b) Some E-literals—like $[\forall x Px \vDash Pa]$—are propositionally justified independently of theory choice and outside the context of an entire logical theory. (cf. §2.2)

Now, would the truth of (b) be problematic for the holist position as it is expressed in (a)? One may suspect that the central argument resulting in (b) is off the mark because a charitable interpretation of the holist position seems able to take on board the whole story of §2.2. After all, the upshot of the argument is that assuming some very general features of deductive entailment, the E-literal $[\forall x Px \vDash Pa]$ will be true under all acceptable entailment-relations, which perhaps doesn't amount to showing that $[\forall x Px \vDash Pa]$ *is justified outside the context of an entire logical theory*, but rather that the E-literal is justified *independently of theory choice* in the sense that no matter what theory you consider it in the context of, it will be justified. Compare, for instance, to a contextualist position about knowledge attributions: the proposition expressed by the claim that '*Subject*, S, *knows that* S *exists*' is not true independently of context in a sense that refutes contextualism, but in the sense that it is true in every context. Thus, on a charitable reading, what the holist claims is that a particular E-literal, like $[\forall x Px \vDash Pa]$, cannot be justified outside the context of a logical theory because a logical theory is what specifies "the bounds of deduction". And so, the holist could accept all the central claims made in §2.2 as part of a broad holistic justification-enterprise.

While this objection completely misses the central point about the booting-property of $[\forall x Px \vDash Pa]$ and how this special feature of the E-literal about universal instantiation gives rise to pre-theoretic justification, let's just assume for the sake of argument that the E-literal $[\forall x Px \vDash Pa]$ doesn't provide us with a direct counterexample to justification holism under a charitable reading of the position. This notwithstanding, the Argument from Pre-Theoretic Universality would pose an indirect challenge to the holistic claim that entire logical theories, not

[23] Thanks to an anonymous reviewer for raising this issue.

individual E-literals, are the primary bearers of justification in the epistemology of logic, i.e., that whatever justification we may have for our individual claims of entailment must be due to the justifiedness of logical theories *en bloc*. Since the propositional justification of foundational E-literals like $[\forall x Px \vDash Pa]$ is orthogonal on the issue of theory choice—illustrated by the argument in §2.2—we could just as well have the opposite order of dependence: whatever justification we have for our logical theories must be due to the basic justifiedness of certain foundational E-literals. It's plainly arbitrary to say that logical theories rather than foundational E-literals are primary without further argument at this point. In fact, at least one of the abductivist virtues, viz., simplicity, seems to support the primacy of a very limited set of foundational E-literals.

This reply can even be strengthened if we notice that not everything hinges on the success of the argument in §2.2 as there are plausible candidates of foundational E-literals other than $[\forall x Px \vDash Pa]$. Consider for instance the E-literal about the inference rule *uniform substitution* instead of universal instantiation. In the end—on the pain of nihilism about deductive entailment—certain entailments need to go through no matter our theoretical differences because giving up on them would mean giving up on deduction as such. Foundational E-literals, like the ones suggested in the present chapter, should come across as a very suitable basis of justification in the epistemology of logic, or at least they should be on par with entire logical theories in this respect.[24]

Circularity

Another objection to the Argument from Pre-Theoretic Universality is that while the proclaimed aim of the argument was to establish $[\forall x Px \vDash Pa]$ as a foundational E-literal of deduction, it ended up merely presupposing the truth of $[\forall x Px \vDash Pa]$.

To unpack this objection a bit, consider the following pattern of reasoning. Suppose that a deductive entailment is valid when all cases where all its premises are true also make its conclusion true. If so, entailment—semantically understood—is essentially tied up with universal quantification over cases. And thus, if the E-literal $[\forall x Px \vDash Pa]$ is true, we get that from '*In all cases where all premises of a valid entailment are true, its conclusion is true*' it follows that '*If this particular model, $\mathfrak{M}$, makes all the premises of a valid entailment true, $\mathfrak{M}$ also makes its conclusion true*'. But how can the fact that this latter claim follows justify the E-literal for UI itself? Or, in other words, how does this fact "ground" the truth of the E-literal $[\forall x Px \vDash Pa]$ in a non-holistic way rather than simply presupposing it?

[24] Further, it has been argued that abduction cannot serve as a neutral arbiter in foundational disputes about logic since in order to use abduction one must first point out the relevant data to assess, and which data is found relevant is not independent of one's foundational views regarding many of the disputes one may hope to solve via abduction (Hlobil, 2021).

In response to this, one should simply bite the bullet and observe that while there was undeniably some circularity involved in establishing the foundational truth of $[\forall x Px \vDash Pa]$, this was both expected and unproblematic from an atomistic perspective. Indeed, the relevant kind of circularity was already highlighted in §2.1 under the label 'Universality-Booting' as a special fact about $[\forall x Px \vDash Pa]$. What makes $[\forall x Px \vDash Pa]$, and perhaps a few other E-literals, stand out from the rest as a foundation of deduction is at least partly their bootstrapping nature, so the relevant kind of circularity is a distinguishing feature of foundational E-literals rather than a bug in the main argument.[25]

Truth-Aptness

Yet another objection to the result from §2.2 is that if UI is definitional with respect to the universal quantifier, then UI is not truth-apt, i.e., the Argument from Pre-Theoretic Universality involves a certain category mistake.

In response, one should notice, yet again, that the argument concerns the E-literal *about* UI, i.e., $[\forall x Px \vDash Pa]$, not the rule UI. In other words, it concerns the claim *that* [UI is valid], or *that* $[\forall x Px$ entails $Pa]$. As $[\forall x Px \vDash Pa]$ is truth-apt, the argument clearly doesn't fall prey to the suggested category mistake.

Free Logic

A final obvious worry is based on the fact that UI fails in standard theories of free logic (Williamson, 1999; Sider, 2010; Nolt, 2021). From this it can be argued that something must be wrong with the argument in §2.2 since $[\forall x Px \vDash Pa]$ cannot be a foundational E-literal of deductive entailment if it fails in logical theories like the standard ones of free logic. Let's spell out the details of this objection.

On standard semantic accounts, the proponent of a free logic has two alternatives. On the one hand, a model of free entailment might allow for two disjoint domains D and D^*, where D is an "inner" domain, which on the standard interpretation consists of existing objects and is the domain of quantification, while D^* is an "outer" domain, usually thought to consist of non-existing objects like, say, Big Foot, Pegasus, the golden mountain etc. While either domain can be empty, their union must be non-empty (by definition). In such models, it's possible for $D \cup D^*$ to be larger than the domain of quantification, and thus $[\forall x Px \vDash Pa]$ could be false. Suppose, for instance, that model $\mathfrak{M}$ is specified such that $D = \{x : x\, is\, human\}$

[25] Note also the literature on the more or less related topics—mentioned in the preamble to this chapter—e.g., the *Adoption Problem* (Carrol, 1895; Kripke, 1974; Berger, 2011; Padro, 2015; Besson, 2019; Cohnitz and Estrada-González, 2019; Finn, 2019; Williamson, 202X); the *Background Logic Problem* (Martin, 2021a,b), and *Hinge Propositions* (Wittgenstein, 1969b; Wright, 2004a,b; Coliva and Moyal-Sharrock, 2016; Ranalli, 2020).

and the symbol 'P' refers to the property of being human. Here, the proposition expressed by the sentence '$\forall xPx$' is true in $\mathfrak{M}$. But suppose then that $D^* = \{Pegasus\}$. This would make $[\forall xPx \vDash Pa]$ false in $\mathfrak{M}$ since the name 'a' could denote Pegasus, who is not human. On the other hand, the proponent of free logic could make do with models that only include the usual domain D (of existing objects), while at the same time allowing for D to be empty and with the interpretation function being partial (leaving the interpretation of some names undefined).

To get our reply going, let's first make the following observation. Free logicians reject UI as we have understood it above and replace it with their own UI-principle based on their preferred logical analysis. In some cases, their analysis would involve an extra clause stating that 'object a exists' (perhaps using an existence predicate denoted '$E!$'). So, as a statement of UI, instead of having $[\forall xPx, \Gamma \vDash Pa]$ with Γ empty, they may have something like $[\forall xPx, E!a \vDash Pa]$. These are two completely general, not relativized, rival principles of universal instantiation, which makes the tension between them a genuine case of *logical disagreement* (Williamson, 1988; Hattiangadi, 2018; Andersen, 2020, 2023b; Hjortland, 2022; Rossi, 2023). Some free logicians may accept that $[\forall xPx \vDash Pa]$ in case 'a' is not an empty name, but reject that this is the (correct) principle of universal instantiation, and endorse $[\forall xPx, E!a \vDash Pa]$ instead. We can make an analogy to the famous case of double negation elimination ('DNE'). It may be that the intuitionist accepts DNE for a limited number of cases that one can specify as an extra clause added to the original DNE-principle, but that doesn't mean they accept DNE; they still reject it.[26]

Nonetheless we don't need to launch anything like a campaign against the legitimacy of theories of free logic *tout court* in order to steer clear of the objection. Even the free logician would accept that, in semantics, Universality is a necessary property of every acceptable deductive entailment-relation, i.e., any modeling of the concept validity must involve universal quantification over cases; and this is all the agreement needed to get the Argument from Pre-Theoretic Universality off the ground. A friend of free logic can thus run the whole story from §2.2 with a version of UI they accept (based on their favored logical analysis). This will not change the brute fact that their preferred logical theory—whatever it may be—is booting in a state of universality.[27]

[26] Thanks to an anonymous reviewer for pressing this point.

[27] A similar reply goes against other theories of logic in which UI fails, e.g., certain theories of quantified modal logic. Such theories are notoriously controversial, however, and it is way beyond the scope of the present chapter to dive into this intricate debate.

3 Conclusion and Corollaries

Let's take stock. We have an argument against justification holism, which is taken to be a necessary component of an abductivist approach to the epistemology of logic. Justification holism asserts that claims of logical consequence can only be justified in the context of an entire logical theory. According to holism, claims of logical entailment cannot be atomistically justified as isolated statements, independently of theory choice. Yet the Argument from Pre-Theoretic Universality has shown that the E-literal about UI is special, i.e., $[\forall x Px \vDash Pa]$ is true under any acceptable deductive entailment-relation, and that giving up on it would amount to giving up on deduction altogether. This makes $[\forall x Px \vDash Pa]$ a foundational E-sentence of deductive entailment, meaning that its propositional justification is independent of theory choice, and its justifiedness emerges from outside the context of a whole logical theory. Ergo: justification holism is false.

Now, given that the argument holds, one may wonder about the collateral consequences, i.e., the wider consequences of the result *vis-à-vis* the epistemology of logic. One (admittedly sketchy) way to assess the corollaries of the result is by considering these three conditionals:

(i) Justification holism entails the falsity of justification atomism.

(ii) Abductivism entails justification holism.

(iii) Anti-exceptionalism about logic entails abductivism.

The truth-values of (i)-(iii) will determine how many applications of *Modus Tollens* ('MT') we can use and what exactly the corollaries will be.

At this point, assessing (i) is smooth sailing. Given the Argument from Pre-Theoretic Universality and that justification atomism is simply the negation of justification holism, we get that justification holism is false by one application of MT. This is the main result of the chapter.

From that result plus the truth of (ii), we could get that abductivism is also false (by MT). However, some might hesitate to accept (ii) as true, e.g., Woods (2019). For while there seems to be an implicit assumption among some abductivists that something akin to Rationalism (Bealer, 1998; BonJour, 1998) and Semanticism (Carnap, 2014; Ayer, 1952) are the only two routes to justification for the atomist, this assumption could—and should—be challenged: why couldn't one have point-wise abductive justification for each axiom of a given logical theory, i.e., why couldn't one justify the axioms point-wise if each one of them explains a distinct set of (empirical) data or evidence? It's not clear why the atomist needs to resort to *a priori* intuitions about logical laws, or proper understanding of the meaning of logical connectives, in order to gain justification. And thus, it's not clear why an endorsement of abductivism forces on us a commitment to holism.

Yet, let's suppose that (ii) is true, then, by MT, we get that abductivism is false. From this result, and the assumption that (iii) is true, we would have the interesting conclusion that anti-exceptionalism about logic is false as well (by MT). However, as the expression 'anti-exceptionalism about logic' is an umbrella term covering different metaphysical, epistemic, methodological, and normative aspects in the philosophy of logic (Martin and Hjortland, 2022), the inference from the negation of abductivism to the negation of anti-exceptionalism should be restricted in order to gain plausibility. Perhaps the inference is most plausible given a certain *epistemic* version of anti-exceptionalism about logic.

Chapter 5

Third Preamble

1 Preamble

At this stage of the monograph we have already met both the Ad Hoc Reading and the Theory Choice Reading of 'logical disagreement'. Next, we'll turn to the Akrasia Reading. In §§1.1-1.4 the reader will find some useful background information setting the stage for Chapter 6 on *logical akrasia*.

1.1 'Logical Disagreement'—The Akrasia Reading

The Greek word 'akrasia' translates literally as 'lack of self-control', but has come to be used as a general term for a weakness of will, i.e., a disposition to act contrary to one's own considered judgement. As is well known, akrasia has interested philosophers ever since antiquity (see, e.g., Plato's *Protagoras* 351a-358d). A given subject S is in an akratic state if and only if (i) S believes that they ought to do action φ, and yet (ii) S does not-φ.

Two clarifying remarks are in order here. First, we read the term 'ought' as *the best way to satisfy S's undefeated desire, all things considered*; where 'undefeated desire' is a desire that prevails over whichever other desires S might have. And we read 'all things considered' in the Davidsonian way, i.e., S didn't fail to take into account any relevant reason (Davidson, 2001). Second, a state of akrasia is not a state of straightforward contradiction. There is no contradiction between S's believing that they ought to do φ and S's doing not-φ, but even so, it should be clear that there is something self-undermining and seemingly incoherent about cases of akrasia. Let's entertain an example:

> **Fred the Philosopher**. Fred is a philosophy PhD student who has a strong

> desire to become a professional philosopher. This, however, is not Fred's only concern—he also wants to earn a bit of money and have a fairly stable lifestyle without too much stress and uncertainty. After careful deliberations with his partner, friends, and family, Fred comes to believe that he ought to apply for a job in software engineering rather than philosophy. Yet, when the time comes, Fred finds himself applying for jobs only in philosophy, none in software engineering.

It will come as no surprise to the reader that the kind of inability to act as one thinks right, which is exemplified by Fred's case, has interested philosophers—and ethicists in particular—ever since antiquity.

More surprising (perhaps) is the vast amount of attention that the analogous phenomenon *epistemic akrasia* has received lately.[1] A driving force behind this interest is the appealing thought that epistemic rationality requires coherence between: (A) an agent's doxastic attitudes in general, and (B) their specific beliefs about what doxastic attitudes are rational.[2]

To illustrate, consider the following case:

> **Anandi the Medical Doctor**. Anandi is a medical resident who correctly figures that dosage $\langle p \rangle$ is appropriate for her patient; and thus believes that $\langle p \rangle$. Suppose she then learns she's been drugged herself, and further that the effects of the relevant drug very often lead to cognitive errors that are hard to detect from the inside. As a result, suppose she believes that ⟨*my belief that p is irrational*⟩ but that she maintains her belief that $\langle p \rangle$ nonetheless.

Other things being equal, Anandi's doxastic state should strike us as irrational because she believes against her own standards of rationality, or as recent epistemological parlance will have it; because she believes *akratically*.

Another similar case will help us grasp a popular *evidentialist* formulation of epistemic akrasia. According to this formulation a subject is epistemically akratic when they are highly confident that proposition $\langle p \rangle$ is true while also believing that the higher-order proposition expressed by ⟨*my current evidence doesn't support p*⟩ is the case. So, if Anna believes that it's going to rain tomorrow while also believing that her evidence at the time doesn't support this,

[1] One of the earliest discussions of epistemic akrasia can be found in (Rorty, 1983). For more recent discussions of the topic, see (Adler, 2002; Owens, 2002; Ribeiro, 2011; Williamson, 2011a; Smithies, 2012; Greco, 2014; Horowitz, 2014; Lasonen-Aarnio, 2014; Williamson, 2014; Sliwa and Horowitz, 2015; Titelbaum, 2015; Roush, 2017; Brown, 2018; Littlejohn, 2018; Worsnip, 2018; Daoust, 2019; Kappel, 2019b; Skipper, 2019; Titelbaum, 2019; Kearl, 2020; Lasonen-Aarnio, 2020; Chislenko, 2021; Skipper, 2021; Christensen, 2021, 2022; Jackson and Tan, 2022; Kauss, 2023).

[2] Examples of *doxasitic attitudes* are: belief-tokens, credences, opinions, judgements etc.

then Anna is in a state of epistemic akrasia. *Prima facie*—at least—Anna's position should strike us as irrational because believing against what one takes one's evidence to support just seems epistemically bad; if not outright paradoxical.

It's examples like Anna's weather forecast and Anandi's drug case that have led some epistemologists to argue for a general anti-akrasia constraint on epistemic rationality (Feldman, 2005b; Smithies, 2012; Titelbaum, 2015; Littlejohn, 2018):

> *The Akratic Principle.* No [epistemic] situation rationally permits any overall [doxastic] state containing both an attitude A and the belief that A is rationally forbidden in one's situation. (Titelbaum, 2019, p. 227)[3] [4]

In spite of various points of ambiguity, this principle is usually taken to imply that you should either have the attitudes you believe you ought to have, or stop believing that you ought to have those attitudes. Hence, in the name of rationality, the Akratic Principle forbids you to have certain combinations of attitudes such as *not* believing that $\langle p \rangle$ while believing that $\langle$*believing p is rationally required*$\rangle$; or having $credence(p) = 0.9$ while believing that $\langle$*having* $credence(p) = 0.9$ *is rationally forbidden*$\rangle$.

Now, while epistemic akrasia is interesting in its own right, it will not be our main concern in Chapter 6. Our primary focus will be on an analogous phenomenon in (formal) logic, viz., *logical akrasia*. One aim of the sixth chapter is to connect the discussion of epistemic akrasia from mainstream epistemology with another existing discussion in the philosophy of logic, which concerns the use of classical logic to prove metatheoretic results (such as soundness and completeness) about a weaker, non-classical logic.[5]

As a rough starting point—for the Akrasia Reading of 'logical disagreement'—we'll simply take logical akrasia to consist in a mismatch between the deductive strength of the background logic one uses to prove metatheoretic results and the logical theory one prefers (officially), i.e., a form of internal incoherence (or intra-personal disagreement if you like) in

[3] Note that the Akratic Principle is sometimes referred to as the *Enkratic Principle* instead, see, e.g., (Skipper, 2019; Field, 2019, 2021).

[4] To get a clearer grasp of Titelbaum's use of the term 'rational'—as displayed in the Akratic Principle—the reader should consult the first appendix of the present monograph, i.e., Chapter 9. Among many other useful details the relevant appendix will give precise definitions of the notions *rational permission* and *rational requirement*.

[5] Common forms of soundness and completeness can be stated as follows. Soundness: $\Gamma \vdash \varphi \Rightarrow \Gamma \models \varphi$. "Everything provable is valid." Completeness: $\Gamma \models \varphi \Rightarrow \Gamma \vdash \varphi$. "Everything valid is provable." Or, as Restall and Standefer (2023, p. 91) put it:

> "Soundness" is a kind of consistency criterion. We don't have both a proof and a counterexample for a single argument. "Completeness" is the opposite. For every argument, we have *either* a proof *or* a counterexample. Soundness and completeness... are the claims that our notions of proof and counterexample are mutually exclusive and exhaustive.

logical theorizing akin to what we saw in the case of epistemic akrasia.[6] In other words, logical akrasia will occur when one explicitly appeals to, or at least implicitly commits to, a logical principle which is not endorsed by one's own theory.

In Chapter 6 we'll be given a more rigorous definition of logical akrasia (and of logical theories), but for now an easy example will suffice to guide our intuitions:

> **Graham the Dialetheist**. Suppose that Graham is a logician who believes that some contradictions are true. He officially prefers the logical theory LP, i.e., the Logic of Paradox. Nevertheless, when doing a completeness proof for LP, Graham finds himself repeatedly appealing to the logical principle Modus Ponens, which is invalid (or only "quasi-valid") in LP. Hence, in his metatheoretic pursuits, Graham appeals to a logical principle which is not endorsed by his own theory.

Here, Graham is in a state of logical akrasia. The case of the dialetheist who—when doing the metatheory of their paraconsistent logic—finds themself using principles that are merely classically valid, is an illustrative example of logical akrasia as it involves a clear incoherence of the kind we are interested in. This dialetheist happens to presuppose logical principles, when producing metatheoretic proofs, that are not endorsed by their own (paraconsistent) standards, and as was the case with the epistemic counterpart, logical akrasia seems self-undermining and irrational. There is just something seemingly hypocritical and problematic about taking *logical validity* to obey a logic weaker than classical, and then continuously developing one's theory of that logic using inferences that are merely classically valid. Or, to put this point more vividly: the paraconsistentist searching for an acceptable metatheory using a classical background logic seems akin to fixing a leaky roof by accustoming oneself to a wet floor.

1.2 From Logic to Epistemology (via Bridge Principles?)

In Chapter 1 we saw that we can't simply take for granted that logic is normative for reasoning. The work of Gilbert Harman (1984; 1986) drives a wedge in between deductive logic and the norms of reasoning, and exposes the need for *bridge principles*. These principles are called for in order to bridge the gap between pure logic and the normative constraints that logic allegedly imposes on our reasoning (MacFarlane, 2004). As we also learned in Chapter 1, bridge principles are roughly of the form:

Bridge. If $\delta(\Gamma \vDash \varphi)$ then $D(\alpha(\Gamma), \beta(\varphi))$,

[6] Logic L_i is *deductively stronger* than logic L_j whenever L_i can prove more, i.e., for every set of well-formed sentences, Γ, the deductive closure of Γ under L_j is a proper subset of the closure of Γ under L_i.

where δ is a doxastic attitude (judging, believing etc.) *vis-à-vis* the entailment $\Gamma \vDash \varphi$ (note that δ can be empty in some cases).[7] D is a deontic operator (varying in scope) constraining the (possibly distinct) doxastic attitudes α, β *vis-à-vis* Γ, and φ, respectively. A few examples of bridge principles are:

- If $\langle \Gamma \vDash \varphi \rangle$, then subject S ought to believe $\langle \varphi \rangle$ if believing every member of Γ.
- If subject S knows $\langle \Gamma \vDash \varphi \rangle$, then S has an *all-things-considered* reason to believe $\langle \varphi \rangle$ if knowing each member of Γ.
- If subject S believes $\langle \Gamma \vDash \varphi \rangle$, then it is permissible for S to believe $\langle \varphi \rangle$ if believing every member of Γ.

In Chapter 6 we'll see that cases of logical akrasia aren't violations of the Akratic Principle in any straightforward way (cf. Chapter 6, §3). The standard of error (the epistemic normativity) of logical akrasia does not in any obvious way arise from logical theorizing alone. One way to overcome this apparent gap—and make logical akrasia immediately interesting to epistemologists—would go via bridge principles of the general format we have just spelled out, but in Chapter 6 we'll take a somewhat different route. The standard of error we'll be concerned with, when considering cases of logical akrasia below, is going to be what Cohnitz and Estrada-González (2019, p. 137) call the "prima facie epistemology of logic", i.e., the epistemic ideal of *Reflective Equilibrium*.[8]

1.3 Reflective Equilibrium

Roughly speaking Reflective Equilibrium ('RE') obtains when there is a balance between our considered judgements (or 'intuitions' as Rawls would say (1951; 2020)) and the general principles that guide them. One the one hand, RE can be viewed as a *philosophical method*, where

[7] To avoid making the formal notation of bridge principles any more clumsy, we simply take for granted that doxastic attitudes are to be had by cognitive agents (rather than indexing the symbols referring to doxastic attitudes to such agents). Note also that while we follow Steinberger (2019a, p. 312) in using the above formalism for generalized bridge principles, the notation is actually somewhat confusing, e.g., the operator D can vary in scope, but still it certainly looks as if it takes a wide scope in the formalism.

[8] It's also worth flagging how our discussion of logical akrasia in Chapter 6 relates to some issues in the logical pluralism debate, especially the normative status of logic and the so-called '*Collapse Problem*' (see, e.g., (Beall and Restall, 2006; Read, 2006; Russell, 2020; Tajer, 2022a)). Roughly, a logical pluralist argues that there can be more than one correct logic (say L_1 and L_2), but consider then a case where we have a set of true (and known) premises Γ, such that a particular conclusion φ follows from Γ in L_1, but not in L_2. If we take logical validity to be truth-preserving—and assume logic to have normative force—then it seems immediately plausible that the pluralist ought to believe the conclusion φ (as a consequence of L_1). But then the pluralist thesis also seems to collapse into monism, viz., monism about L_1; since L_1 allows us more true beliefs. This problem is sometimes referred to as the 'upward collapse': collapse into the strongest consequence relation (but there is also, at least in principle, a problem of downward collapse, i.e., into the weakest logic by way of "modesty" in a certain sense).

RE consists in working back and forth among our considered judgements about particular cases and the principles that govern them; and revising any of these elements wherever necessary in order to achieve an acceptable coherence, or balance. On the other hand, RE can be seen as an epistemically desirable *doxastic state*, viz., the output applying the RE-method properly such that the resulting state is a "provisional fixed point" (Daniels, 1979, p. 267). A simple example of an RE-state from normative ethics could be a perfect balance between the considered judgement that (a) *I ought help the homeless man in front of Tesco on Market Street* and the general principle that (b) *One ought to help the homeless*.

Further, Norman Daniels (1979; 1996) claims that *Wide Reflective Equilibrium* ('WRE') consists in an ordered triple of sets: (A) a set of considered judgements; (B) a set of general principles; and (C) a set of relevant background theories. Balancing our judgements about particular cases against our general principles only gives us *narrow* reflective equilibrium, while consistency with (C) takes us all the way to WRE. Without the involvement of (C) one would allegedly run the risk of our principles of (B) being merely "accidental generalizations" rather than genuine "objective laws."

In the context of logic, Michael Resnik (1985; 1996; 2004) identifies RE between one's logical theory and considered judgements about *logicality* (i.e., validity, consistency, implication, equivalence etc.) whenever:

> ...the theory rejects no argument that one is determined to preserve and countenances no argument that one is determined to reject... (Resnik, 1996, p. 493)

Thus we have at least two interpretations of reflective equilibrium, depending on how we read the term 'reject': (1) one's theory judges valid every argument one is determined to preserve; and (2) one's theory doesn't judge invalid any argument one is determined to preserve. In our discussion of logical akrasia in Chapter 6 we'll see that especially interpretation (1)—i.e., the *strong* interpretation—of reflective equilibrium will be relevant to us.[9]

1.4 Peano Arithmetic in the Post-Gödel Era

In Chapter 6 we'll also be interested in a logical theory of *Peano Arithmetic* ('PA'), i.e., the standard logic of the natural numbers if you like. In the present subsection we'll sketch some core definitions of PA, following the full presentations in (Berto, 2011; Cieśliński, 2017).

The first definition specifies the language of first-order arithmetic (henceforth denoted '$\mathfrak{L}_{PA}$'):

> $\mathfrak{L}_{PA}$: The language of first-order arithmetic contains the usual logical vocabulary (connectives, quantifiers etc.) and auxiliary symbols such as brackets and

[9] See (Woods, 2019) for an interesting argument against the ideal of RE in logical theorizing.

punctuation marks. The set of primitive extralogical symbols is $\{+, \times, 0, S\}$ denoting addition, multiplication, zero, and the successor function, respectively.

Terms, formulas, and sentences, of $\mathfrak{L}_{PA}$ are also defined in the usual way. In particular, sentences of $\mathfrak{L}_{PA}$ are defined as formulas without any free variables. So, $\forall x(x + 0 = x)$ is a sentence of $\mathfrak{L}_{PA}$, while $\exists x(S(y))$ is not.

The formalized axioms of PA are:

1. $\forall x(S(x) \neq 0)$
2. $\forall x \forall y(S(x) = S(y) \rightarrow x = y)$
3. $\forall x(x + 0 = x)$
4. $\forall x \forall y(x + S(y) = S(x + y))$
5. $\forall x(x \times 0 = 0)$
6. $\forall x \forall y(x \times S(y) = (x \times y) + x)$
7. $\{[\Phi(0) \wedge \forall x(\Phi(x) \rightarrow \Phi(S(x)))] \rightarrow \forall x(\Phi(x)) : \Phi(x) \in \mathfrak{L}_{PA}\}$

Notice that (7) is the set of arithmetical sentences falling under the *axiom schema* of mathematical induction, i.e., it's an infinite set of axioms rather than just a single axiom.

Note also that since $\mathfrak{L}_{PA}$ doesn't contain any primitive numerals besides 0, numerals of $\mathfrak{L}_{PA}$ are in general specified as terms of the following form: $S \ldots S(0)$, i.e., terms obtained by preceding the symbol 0 with arbitrarily many successor symbols.

Obviously $\mathfrak{L}_{PA}$ allows us to express claims about the natural numbers in the theory PA, e.g., claims concerning addition and multiplication etc. Otherwise the theory wouldn't really be worth its salt. But more important for our purposes below is that we'll tacitly assume some form of *coding* (or Gödel-numbering) throughout Chapter 6. As Kurt Gödel (1931) showed it is possible to define a procedure, starting with assigning natural numbers to primitive expressions of $\mathfrak{L}_{PA}$, and then extending the assignment to more complex syntactical objects. Eventually unique numbers become assigned to terms, formulas, and sequences of formulas; and it effect, we can then view some statements of first-order arithmetic as assertions about syntax. In other words, it becomes possible for us to use PA "introspectively", i.e., in making assertions about the theory PA itself. The most famous example of this is of course the *Gödel Sentence* ('G'), which is at the heart of Gödel's Second Incompleteness Theorem. The sentence G states about itself (via such-and-such substitution operations) that it isn't a provable sentence in PA (Berto, 2011, p. 92)

While it isn't essential to us *how* the encoding from linguistic expressions to numbers is done—Gödel exploited the Unique Prime-Factorization Theorem to this end—it's important to note that it *can* be done.[10] For in Chapter 6 we'll need to appeal to Gödel's Second Incompleteness Theorem in order to state the Dilemma of Logical Akrasia. The incompleteness result involves the consistency claim—Con_{PA}—stating that the logical theory PA is consistent, which in a way is just a regular claim made in $\mathfrak{L}_{PA}$, and yet, this is only the case *indirectly* via our tacit coding procedure.

[10] Gödel's original coding employs prime factorization: a finite sequence of numbers $n_1 \ldots n_k$ will be coded by the number $2^{(n_1+1)} \times 3^{(n_2+1)} \times \ldots \times p_k^{n_k+1}$, where p_k is the k-th prime.

Chapter 6

Logical Akrasia

The aim of this chapter is twofold. First, §1 and §2 introduce the novel concept *logical akrasia* by analogy to epistemic akrasia. If successful, the initial sections will draw attention to an interesting akratic phenomenon which has not received much attention in the literature on akrasia (although it has been discussed by logicians in different terms). Second, §3 and §4 present a dilemma based on logical akrasia. From a case involving the consistency of Peano Arithmetic and Gödel's Second Incompleteness Theorem it's shown that either we must be agnostic about the consistency of Peano Arithmetic or akratic in our logical theorizing. If successful, these sections will underscore the pertinence and persistence of akrasia in logic (by appeal to Gödel's seminal work). §5 concludes with a brief epilogue suggesting a way of translating the dilemma of logical akrasia into a case of regular epistemic akrasia; and further how one might try to escape the dilemma when it's framed this way.

Keywords

Epistemic Akrasia; Logical Akrasia; Epistemic Rationality; Logical Theories; Gödel's Incompleteness Theorem; The Dilemma of Logical Akrasia

1 Prologue

The Greek word 'akrasia' translates literally as 'lack of self-control', but has come to be used as a general term for a *weakness of will*, i.e., a disposition to act contrary to one's own considered judgement. It will come as no surprise that such inability to act as one thinks right has interested ethicists since antiquity.

More surprising (perhaps) is the vast amount of attention that the analogous phenomenon *epistemic akrasia* has received lately.[1] A driving force behind this interest is the appealing thought that epistemic rationality requires coherence between: (A) an agent's doxastic attitudes in general, and (B) their specific beliefs about what doxastic attitudes are rational.[2] To illustrate, consider the medical resident Anandi who correctly figures that dosage $\langle p \rangle$ is appropriate for her patient; and thus believes that $\langle p \rangle$. Suppose she then learns she's been drugged herself, and further that the effects of the relevant drug very often lead to cognitive errors that are hard to detect from the inside. As a result, suppose she believes that ⟨*my belief that p is irrational*⟩ but that she maintains her belief that $\langle p \rangle$ nonetheless. Other things being equal, Anandi's doxastic state should strike us as irrational because she believes against her own standards of rationality, or as recent epistemological parlance will have it; because she believes *akratically*.[3]

Examples like Anandi's drug case have led some epistemologists to argue for a general anti-akrasia constraint on epistemic rationality (Feldman, 2005b; Smithies, 2012; Titelbaum, 2015; Littlejohn, 2018):

[1] See (Adler, 2002; Owens, 2002; Ribeiro, 2011; Williamson, 2011a; Smithies, 2012; Greco, 2014; Horowitz, 2014; Lasonen-Aarnio, 2014; Williamson, 2014; Sliwa and Horowitz, 2015; Titelbaum, 2015; Roush, 2017; Brown, 2018; Littlejohn, 2018; Worsnip, 2018; Daoust, 2019; Kappel, 2019b; Skipper, 2019; Titelbaum, 2019; Kearl, 2020; Lasonen-Aarnio, 2020; Chislenko, 2021; Skipper, 2021; Christensen, 2021; Jackson and Tan, 2022; Horowitz, 2022; Kauss, 2023).

[2] Examples of *doxasitic attitudes*: belief-tokens, credences, opinions, judgements etc.

[3] According to a popular *evidentialist* formulation a subject is epistemically akratic when they are highly confident that proposition $\langle p \rangle$ is true while also believing that the higher-order proposition expressed by ⟨*my current evidence doesn't support p*⟩ is the case. So, if Anna believes that it's going to rain tomorrow while also believing that her evidence at the time doesn't support this, then Anna is in a state of epistemic akrasia. Prima facie—at least—Anna's overall doxastic state should strike us as irrational. Since believing against what one takes one's evidence to support just seems epistemically bad; if not outright paradoxical. To this end, the card-carrying evidentialist Richard Feldman wonders "*...what circumstances could make [epistemic akrasia] reasonable...*" (Feldman, 2005b, p. 109).

> *The Akratic Principle.* No [epistemic] situation rationally permits any overall [doxastic] state containing both an attitude A and the belief that A is rationally forbidden in one's situation. (Titelbaum, 2019, p. 227)[4] [5]

This principle is taken to imply that you should either have the attitudes you believe you ought to have, or stop believing that you ought to have those attitudes. Hence, in the name of rationality, the Akratic Principle forbids you to have certain combinations of attitudes such as *not* believing that $\langle p \rangle$ while believing that $\langle$*believing p is rationally required in one's situation*$\rangle$; or having $credence(p) = 0.9$ while believing that $\langle$*having* $credence(p) = 0.9$ *is rationally forbidden in one's situation*$\rangle$.

We'll return to the Akratic Principle in due course (cf. §3), but for now let's consider a widely discussed case from the literature on epistemic akrasia to further guide our intuitions. The case concerns a sleep deprived detective, Sam, who possesses misleading higher-order evidence (i.e., misleading evidence about what his first-order evidence supports):

> **Sleepy Detective**. Sam is a police detective, working to identify a jewel thief. He knows he has good evidence—out of the many suspects, it will strongly support one of them. Late one night, after hours of cracking codes and scrutinizing photographs and letters, he finally comes to the conclusion that the thief was Lucy. Sam is quite confident that his evidence points to Lucy's guilt, and he is quite confident that Lucy committed the crime. In fact, he has accommodated his evidence correctly, and his beliefs are justified. He calls his partner, Alex. "I've gone through all the evidence," Sam says, "and it all points to one person! I've found the thief!" But Alex is unimpressed. She replies: "I can tell you've been up all night working on this. Nine times out of the last ten, your late-night reasoning has been quite sloppy. You're always very confident that you've found the culprit, but you're almost always wrong about what the evidence supports. So your evidence probably doesn't support Lucy in this case." Though Sam hadn't attended to his track record before, he rationally trusts Alex and believes that she is right—that he is usually wrong about what the evidence supports on occasions similar to this one. (Horowitz, 2014, p. 719)

Provided the information of Sleepy Detective—and the background assumption that respecting one's *total* evidence is an important standard of epistemic rationality—what is the rational response to Sam's predicament? In other words, what doxastic attitude should he

[4] Note that the Akratic Principle is sometimes referred to as the *Enkratic Principle* instead, see e.g., (Skipper, 2019; Field, 2019, 2021).

[5] For further details on Titelbaum's use of the term 'rational' consult (Titelbaum, 2015, 2019; Skipper, 2019) and Appendix 1 of the present monograph. See also (Bradley, 2021; Carr, 2021) for recent discussions of *ideal* versus *non-ideal* epistemic rationality.

hold with respect to the identity of the thief? And what should he believe about what his first-order evidence supports?[6]

The literature is divided into three main camps. According to *Steadfast* views, Sam should simply stick to his guns. That is to say, he should keep both his high confidence that $\langle p \rangle$ (i.e., $\langle$*Lucy is the thief*$\rangle$) and his belief that this is what his first-order evidence supports (Kelly, 2005; Titelbaum, 2015). A reason in favor of this response is that Sam actually got things right to begin with. So even though the later testimony from his partner Alex is higher-order evidence suggesting that his assessment of the first-order evidence is unreliable due to sleep deprivation, this is *in fact* misleading on the particular occasion.

In contrast, *Conciliatory* views hold that Sam should reduce confidence both with respect to proposition $\langle p \rangle$ and the higher-order proposition stating that $\langle$*my first-order evidence supports* $p\rangle$ (Feldman, 2005b; Christensen, 2007).[7] A reason in favor of this position is that from Sam's first-person perspective the higher-order evidence constituted by Alex's testimony seems undefeated. Since Sam rationally trusts Alex to be right about his unreliable track record in relevantly similar circumstances, he should reduce his confidence at both first- and higher-order level.[8]

Notice that although steadfast and conciliatory views disagree about the rational response to cases like Sleepy Detective, they agree that Sam's confidence in $\langle p \rangle$ shouldn't conflict with his belief about what the first-order evidence supports. That is, both camps accept that any such level-incoherence is epistemically irrational.

Level-Splitting views dispute this. According to the level-splitter it can sometimes be epistemically rational to have a high confidence that $\langle p \rangle$ while also believing the higher-order proposition expressed by the sentence $\langle$*my first-order evidence doesn't support* $p\rangle$. Imagine, for instance, a long deductive proof written on a whiteboard, and suppose that Beth thinks through the proof and comes to rationally believe a series of claims from which she competently deduces their conjunction, $\langle p \rangle$.[9] Assume (quite plausibly) that Beth comes to rationally believe $\langle p \rangle$ by these means. Yet Beth knows that people like her—in similar situations involving long deductions—often make inferential errors. So, it may well be highly probable on her higher-order evidence that she has made an inferential error in the current situ-

[6] Sam's *first-order* evidence includes (propositions about) the letters and photographs that he was looking through as he worked late at night.

[7] One should realize that while a higher-order proposition like $\langle$*my first-order evidence supports that* $p\rangle$ has *positive* normative force with respect to proposition $\langle p \rangle$, i.e., it might make it rational for you to believe that $\langle p \rangle$; other higher-order propositions like $\langle$*any epistemic situation makes it rationally forbidden to believe that* $p\rangle$ has *negative* normative force with respect to proposition $\langle p \rangle$.

[8] For canonical work on *defeaters* in epistemology the reader should consult (Pollock, 1970, 1974, 1984, 1986, 1994). Note also that Christensen might be said to lean towards a level-splitting rather than conciliationist view about akrasia in his more recent work on the topic.

[9] Denoting a conjunction using the symbol 'p' might be thought to overload the notation, but we allow this here for the sake of simplicity.

ation, which suggests that $\langle$*her first-order evidence doesn't support* $p\rangle$ after all (even though we can stipulate that her belief in the truth of $\langle p\rangle$ is *in fact* correct).[10] To be sure, the intended interpretation here is that the knowledge Beth possesses about people's shortcomings in situations relevantly similar to hers should be taken as higher-order evidence against her first-order attitude towards $\langle p\rangle$, but according to level-splitting views, what goes on at higher-order level need not affect the rationality of Beth's first-order attitudes.[11] Thus—by level-splitting lights—this is a scenario where Beth can have a high confidence in $\langle p\rangle$ while also believing the higher-order proposition expressed by $\langle$*my first-order evidence doesn't support* $p\rangle$ and be rational nonetheless.

As with Beth's logic case, a level-splitting response to Sleepy Detective would have it that Sam should remain highly confident that $\langle$*Lucy is the thief*$\rangle$ and simultaneously believe that this isn't supported by his first-order evidence. Epistemologists such as Williamson (2011a; 2014), Lasonen-Aarnio (2014; 2020), Wedgewood (2012), and Weatherson (2010), have all favored level-splitting views although their reasons for doing so diverge.

2 Logical Akrasia

While epistemic akrasia is interesting in its own right, it will not be our main concern. Our primary focus will be on an analogous phenomenon in (formal) logic. The remaining sections aim to connect the discussion of epistemic akrasia from mainstream epistemology with another existing discussion in the philosophy of logic, which concerns the use of classical logic to prove metatheoretic results (such as soundness and completeness) about a weaker, non-classical logic. It will also be suggested that some level of logical akrasia is unavoidable because of Gödel's Second Incompleteness Theorem.

As a rough starting point we'll take *logical akrasia* to consist in a mismatch between the deductive strength of the background logic one uses to prove metatheoretic results and the logical theory one prefers (officially), i.e., a form of level-incoherence in logical theorizing akin to what we saw in the case of epistemic akrasia.[12] So, in other words, logical akrasia will occur when one explicitly appeals to (or at least implicitly commits to) a logical principle which is not endorsed by one's own theory.[13] At this point we won't distinguish between *meta-logic*

[10] Notice the structural analogy between Beth's logic case above and the well-known *Preface Paradox* (Makinson, 1965; Sorensen, 2020).

[11] For further clarification of the distinction between first-order and higher-order evidence, see (Christensen, 2010; Skipper, 2021).

[12] Logic L_i is *deductively stronger* than logic L_j whenever L_i can prove more, i.e., for every set of well-formed sentences, Γ, the deductive closure of Γ under L_j is a proper subset of the closure of Γ under L_i.

[13] A concrete example of an *implicit* commitment to a logical principle could be committing to the excluded middle via an explicit endorsement of Peirce's Law—i.e., $(((\varphi \rightarrow \psi) \rightarrow \varphi) \rightarrow \varphi)$, where lowercase letters from the Greek alphabet are metavariables. For more on the topic of implicit commitments in logical theorizing,

and *metatheory*, but we'll discuss the potential importance of drawing this distinction in §2.1.

Now, to provide a concrete example of logical akrasia, consider the following passage from Beall and Restall:

> **The Intuitionist**. What we have presented is a straightforward account of Tarski's model theory for classical predicate logic, and a simple account of truth conditions in a possible worlds semantics. We have claimed that such accounts deliver classical logic. Is this indeed the case? It is commonly thought that this is the case, but in present company we may have reason to question this thought. Upon an inspection of the usual soundness and completeness proofs, we shall see that the full power of classical logic is required to complete the proof. To show, for example, that in every model $A \vee \neg A$ is satisfied, we need to show, for each model $\mathfrak{M}$, that $\mathfrak{M} \Vdash A$ or $\mathfrak{M} \nVdash A$. But this is an instance of the excluded middle! An intuitionist (for example) who rejects the law of the excluded middle will not endorse this reasoning. What can we say about this? (Beall and Restall, 2006, p. 39)

One thing we could say—to answer Beall and Restall's query—is that the intuitionist is in a state of logical akrasia. The case of the intuitionist logician who, when doing the metatheory of intuitionistic logic, finds themselves using classical (nonconstructive) principles, is an illustrative example of logical akrasia as it involves a clear incoherence of the kind we are interested in. In sum: this logician happens to presuppose logical principles, when producing metatheoretic proofs, that are not endorsed by their own (intuitionistic) standards, and as was the case with the epistemic counterpart, logical akrasia seems self-undermining and irrational. Or, to put this point more vividly: the intuitionist searching for an acceptable metatheory using a classical background logic seems akin to fixing a leaky roof by accustoming oneself to a wet floor.

2.1 Defining Logical Akrasia

So far, so good! Let's now define logical akrasia in a more regimented fashion:

> *Logical Akrasia*. Subject S, preferring logical theory T, is in a state of logical akrasia if and only if S commits to a logical principle that S's preferred logical theory T fails to endorse as valid.[14]

the reader should consult (Cieśliński, 2017; Horsten and Leigh, 2017; Fischer et al., 2021).

[14] Define a *logical theory* in the standard way. A logical theory is an ordered pair $\mathrm{T} = \langle \mathcal{L}_i, \mathcal{L}_i^M \rangle$ such that $\mathcal{L}_i$ is an object-logic and $\mathcal{L}_i^M$ a metatheory.

As this definition is too coarse grained to capture all relevant cases, we further distinguish between:

Weak S, preferring logical theory T, is weakly akratic if S commits to a logical principle that S's logical theory T fails to endorse as valid;

Strong S, preferring logical theory T, is strongly akratic if S commits to a logical principle that S's logical theory T rejects as *invalid*.

Based on this weak/strong distinction we get two non-equivalent versions of Logical Akrasia. To appreciate this, consider a case where an intuitionist commits to an instance of the excluded middle in a restricted situation, and suppose that their theory doesn't endorse this as valid. Then, the intuitionist can extend their theory at a later stage such that the instance of the excluded middle becomes endorsed as valid (just stipulate that the situation is decidable)—e.g., it happens to be a case concerning a quantifier-free sentence of arithmetic like $2 + 2 = 4$. Before the extension they were being weakly akratic, but not afterwards.[15]

Yet one should acknowledge that the tenability of the strong/weak distinction depends on the possible division: meta-logic/metatheory. There is, for instance, no difference between *failing to endorse as valid* and *rejecting as invalid* if one holds a formal and complete meta-*logic* rather than a non-formal and incomplete meta*theory*.[16] If one's meta-logic is formal and complete over its domain, then the distinction between strong and weak logical akrasia collapses (since in that case everything which is not valid is simply invalid). When, say, an intuitionist holds a formal and complete meta-logic and the law of the excluded middle has counterexamples, then failing to endorse the excluded middle as valid and rejecting it as invalid amount to exactly the same.

On the one hand, we have *object-logic* $\mathcal{L}_i = \langle \varPhi, \varSigma, P \rangle$, where $\varPhi$ specifies a language of both logical and non-logical vocabulary while $\varSigma$ gives a syntax for $\varPhi$ (determining its well-formed formulas). P in turn provides a set of inference rules and/or axioms for syntactic manipulation of the symbolic strings that $\varPhi$ gives rise to.

On the other hand, we find *metatheory* $\mathcal{L}_i^M = \langle \varPhi_M, I, \models_\varPhi, \vdash_\Phi \rangle$, the first element of which specifies a meta-language $\varPhi_M$. The second element I gives a semantics expressed in that meta-language, i.e., it lays down individual truth-conditions for the logical constants of $\varPhi$ using $\varPhi_M$. The third element $\models_\varPhi$ authorizes a semantic consequence relation between well-formed formulas of Φ. Finally, $\vdash_\Phi$ defines a syntactic consequence relation for well-formed formulas in Φ.

[15] Note that insofar as the distinction between weak/strong akrasia is relevant at all it isn't just relevant to the case of intuitionism. In fact it seems relevant to many, perhaps even most, non-classical logicians like Field, Kripke, Ripley, Beall etc. For they all take classical logic to be valid in non-problematic contexts. Some non-classical logicians are more 'hardcore' and go non-classical all the way down (Priest (2006), Weber et al. (2016) etc.), but this isn't the norm.

[16] For more on logic(s) and *formality*, see for example (MacFarlane, 2000; Beall and Restall, 2006; Dutilh Novaes, 2012; Mortensen, 2013).

2.2 Logical Akrasia and Incoherence

As we have seen above, states of logical akrasia seem to be incoherent. A point which is also frequently underscored in the literature:

> If you take 'logically valid' to obey a logic weaker than classical, you shouldn't ultimately be satisfied with developing your theory of that logic using inferences that are merely classically valid... (Field, 2017, p. 14)
>
> ...what a strange approach to take, if one believes logic X is the correct logic. Why use an alien logic for one's metatheory—and if one does, why trust the result? (Read, 2006, p. 208)
>
> ...it would be untoward in a logic to appeal in proof of its adequacy to principles in which the logic in question does not believe. (Meyer, 1985, p. 13)
>
> If he rejects classical logic for the object language, how is he entitled to rely on it for the metalanguage? (Williamson, 2020b, p. 6)

These quotes notwithstanding logical akrasia is deeply entrenched in our contemporary logical theorizing. For it is no secret that classical logic serves as the golden standard in evaluations of non-classical logics (Schurz, 2021), i.e., it's common practice to take classical (first-order) logic as the "neutral" backdrop against which we evaluate non-classical logics. Examples are: Łukasiewicz' three-valued logic, Kleene's (strong) three-valued logic, Brouwer's intuitionistic logic, Priest's paraconsistent logic etc.[17]

Where does this leave us? Is the current modus operandi of non-classical theorizing severely misguided? Well, insofar as we want *reflective equilibrium* (Resnik, 1985, 1996, 2004) between our logical theories and considered judgements about *logicality* (i.e., validity, consistency, implication, equivalence etc.), there is a sense in which the answer is affirmative. According to Michael Resnik, one's preferred logical theory and considered judgements about logicality are in a state of reflective equilibrium when:

> ...the theory rejects no argument that one is determined to preserve and countenances no argument that one is determined to reject... (Resnik, 1996, p. 493)

Thus we have at least two interpretations of reflective equilibrium, depending on how we read the term 'reject': (1) one's theory judges valid every argument one is determined to preserve; and (2) one's theory doesn't judge invalid any argument one is determined to preserve. For weak Logical Akrasia to violate reflective equilibrium, we need interpretation (1); not interpretation (2) (cf. §2.1). Or, as one may otherwise put it, for weak Logical Akrasia to violate

[17] Consult (Priest, 2008) for classical evaluations of each of the non-classical logics mentioned above. See (Bacon, 2013) for a discussion of *non-classical* metatheories for non-classical logics.

reflective equilibrium, we need the strong interpretation of reflective equilibrium, not the weak one. The ideal of strong reflective equilibrium is incompatible with states of weak Logical Akrasia. Hence, the golden standard of non-classical logicians—committing themselves to classical principles in their metatheoretical pursuits—appears to be a standard of fool's gold in at least one sense.[18]

2.3 The Analogy with Epistemic Akrasia

If we return to the analogy between epistemic and logical akrasia for a minute, we can now appreciate how both the weak and strong version of Logical Akrasia resemble the standard definitions of epistemic akrasia in various ways.

Recall first the appealing thought from §1 stating that epistemic rationality requires coherence between: (A) an agent's doxastic attitudes in general, and (B) their specific beliefs about what doxastic attitudes are rational. Epistemic akrasia occurs whenever S adopts doxastic attitudes that don't live up to S's own standards of rationality. Logical akrasia, similarly, occurs when S commits to logical principles that don't live up to S's own standards of logic. So, in both cases the problem is one of not meeting one's own ideal (rather than pursuing a spurious ideal).

Adding further to the analogy, we have no trouble imagining what *Steadfast* and *Conciliatory* responses to logical akrasia would look like. If one asserts that logical akrasia calls for a revision of one's logical commitments or theory, then it would count as a conciliatory view. If one, on the other hand, submits that no revision is required, it would be a steadfast view. One could also cook up a special *Level-Splitting* kind of steadfastness without noteworthy ingenuity. Assume, say, that one is a logical pluralist (of some sort), then one may argue that there are benign cases of logical akrasia. Given the pluralist's dictum that more than one logic can be correct, the level-incoherence of logical akrasia need not be problematic. Alternatively one might be able to pull off a level-splitting response by appealing to logical instrumentalism (Haack, 1974), i.e., the view that it doesn't make sense to think of logics as being correct or incorrect rather they are simply variously useful or not. It seems plausible to suggest that appealing to instrumentalism in some way could dissolve the tension in cases of logical akrasia to the level-splitter's satisfaction.[19] [20]

[18] See (Priest, 2006) for further discussion of this seemingly self-undermining practice of some non-classical logicians. Note also the potentially interesting distinction between sub-classical and contra-classical logics.

[19] A popular version of logical pluralism can be found in (Beall and Restall, 2000, 2006). Consult (Russell, 2019b) for an extensive overview.

[20] Interestingly, Tim Williamson who is a card-carrying level-splitter with respect to epistemic akrasia, doesn't seem to be one when it comes to logical theorizing. See for instance his recent review of Kit Fine's compatibility semantics in Mind (Williamson, 2020b). To be fair, however, as Williamson discusses in his review, the issue with Fine's compatibility semantics is rather complicated and depends on the motivation for adopting the relevant non-classical logic in the first place. If one's motivation is to model reasoning involving vagueness, then appealing

3 Enter Gödel: Rationality and Logical Akrasia

As advertised earlier, the two sections §3 and §4 make use of Gödel's (in)famous Second Incompleteness Theorem (Gödel, 1931) to pose a dilemma based on logical akrasia.

Gödel's theorem establishes that: *assuming Peano Arithmetic (PA) is consistent, PA doesn't derive Con_{PA}*. In other words, if the logical theory of PA is consistent, then PA cannot derive its own consistency.[21]

Consider now the following akratic puzzle:

> **Logical Akrasia and Peano Arithmetic**. In using Peano Arithmetic, PA, subject S is at least implicitly committing to (and relying on) PA's consistency. After all, if the theory were inconsistent it's no help in sorting out truths from falsehoods. But if S's theory is PA, then the theory doesn't itself prove that PA is consistent (by Incompleteness). So, in that case S's theory doesn't prove the claim that S is committed to, and consequently S is at least weakly akratic (cf. §2.1). Of course, S can easily extend S's theory (why not?), and then con-

to classical reasoning in metalogical (and therefore mathematical) proofs may be innocuous, since mathematics is presumably precise.

[21] Here we simply take the logical theory of PA to be classical first-order logic extended with seven axioms. The language of first-order arithmetic can be specified as follows.

> $\mathfrak{L}_{PA}$: The language of first-order arithmetic contains the usual logical vocabulary (connectives, quantifiers etc.) and auxiliary symbols such as brackets and punctuation marks. The set of primitive extralogical symbols is $\{+, \times, 0, S\}$ denoting addition, multiplication, zero, and the successor function, respectively.

Terms, formulas, and sentences, of $\mathfrak{L}_{PA}$ are also defined in the usual way. The formalized axioms of PA are: **(Ax1)** $\forall x(S(x) \neq 0)$; **(Ax2)** $\forall x \forall y(S(x) = S(y) \rightarrow x = y)$; **(Ax3)** $\forall x(x + 0 = x)$; **(Ax4)** $\forall x \forall y(x + S(y) = S(x + y))$; **(Ax5)** $\forall x(x \times 0 = 0)$; **(Ax6)** $\forall x \forall y(x \times S(y) = (x \times y) + x)$; **(Ax7)** $\{[\Phi(0) \wedge \forall x(\Phi(x) \rightarrow \Phi(S(x)))] \rightarrow \forall x(\Phi(x)) : \Phi(x) \in \mathfrak{L}_{PA}\}$. Notice that axiom 7 is really the set of arithmetical sentences falling under the axiom schema of mathematical induction, i.e., it's an infinite set of axioms rather than just a single axiom. Obviously $\mathfrak{L}_{PA}$ allows us to express claims about the natural numbers in the theory PA, e.g., claims concerning addition and multiplication etc. But what is more important for our purposes below is that we'll tacitly assume some form of *coding*. As Kurt Gödel (1931) showed it is possible to define a procedure, starting with assigning natural numbers to primitive expressions of $\mathfrak{L}_{PA}$, and then extending the assignment to more complex syntactical objects. Eventually unique numbers become assigned to terms, formulas, and sequences of formulas; and it effect, we can then view some statements of first-order arithmetic as assertions about syntax. In other words, it becomes possible for us to use PA "introspectively". The most famous example of this is of course the *Gödel sentence* ('G'), which is at the heart of Gödel's Second Incompleteness Theorem. The sentence G states about itself (via such-and-such substitution operations) that it isn't a provable sentence in PA (Berto, 2011, p. 92). While it isn't essential to us *how* the encoding from linguistic expressions to numbers is done—Gödel exploited the Unique Prime-Factorization Theorem to this end—it's important to note that it *can* be done. For below we'll appeal to the consistency claim—Con_{PA}—stating that the logical theory PA is consistent, which in a way is just a regular claim made in $\mathfrak{L}_{PA}$, and yet, this is only the case *indirectly* via our tacit coding procedure.

> sider PA with Con_{PA} added as an axiom. This would provide S with a proof of Con_{PA} in a single line. But alas, now the issue of akrasia arises at the level of appeal to the consistency of that theory, i.e., $PA + Con_{PA}$. And so on ad infinitum...

In S's use of PA it turns out that S is not just committed to the theory PA itself, but also the stronger theory $PA + Con_{PA}$. Ergo, S is committed to a logical principle that their theory cannot prove, and thus S is in a state of logical akrasia.[22] By construction, there is no way for S to avoid committing to Con_{PA} (or $Con_{PA+Con_{PA}}$, or...) and escaping their akratic state (on the pain of triviality). While this doesn't necessarily show that S is irrational when using PA, it does entail that S is committed to something which goes beyond S's own theory in such situations.

Yet this is a *positive* kind of mismatch—i.e., S's background logic can prove something which S's preferred logical theory cannot—rather than a violation of a *negative* rationality constraint such as the previously mentioned Akratic Principle. Recall the Akratic Principle stating that:

> No [epistemic] situation rationally permits any overall [doxastic] state containing both an attitude A and the belief that A is rationally forbidden in one's situation. (cf. §1)

As this is a negative principle in the sense that it involves an assertion about rationally forbidden states, the puzzle of logical akrasia is not in any obvious way a violation of it. Unless we are ready to grant that it's rationally forbidden to commit oneself to something which one cannot prove, of course, but this seems overly strong. Nobody in their right mind would suggest that provability is a plausible guide to rationality *simpliciter*.[23]

[22] Some readers might hesitate to admit that the case results in logical akrasia because they don't see Con_{PA} as a genuine *logical* commitment: *Why isn't* Con_{PA} *considered a further implicit, non-logical claim which* S *commits to?*

The answer is straightforward. Con_{PA} is just a regular claim made in the language of PA (and definitely not a contingent empirical fact). There may of course be a sense in which PA doesn't count as strictly "logical" but rather as "mathematical." Even so the kind of commitment S holds with respect to Con_{PA} is not essentially different from the one S has towards the axioms of PA (though implicit). Hence—upon reflection—there is indeed a certain kind of akrasia (about the logic of the natural numbers if you like) arising in the puzzle of Logical Akrasia and Peano Arithmetic.

[23] Naturally it could be argued that provability is a plausible guide to rationality in a certain *narrow* sense. It's clear enough that we can rationally believe many contingent propositions that we cannot prove to be correct, but in the case above we are not concerned with any old contingent proposition. We are concerned with the proposition that $\langle PA$ is consistent$\rangle$, and it's not immediately clear how it could be rational to believe Con_{PA} given that it cannot be proved using one's logical theory. See for example (Gentzen, 1936; Chow, 2019) for further discussion of this non-trivial question.

Nonetheless the puzzle of logical akrasia does illustrate a clash with our ideals concerning epistemic rationality insofar as reflective equilibrium is among them. As logically akratic states cannot be in reflective equilibrium—given interpretation (1) from §2.2—the case above does indeed suggest that S's doxastic state is epistemically irrational in a certain sense. How bad this sort of irrationality looks to the epistemologist does of course depend on the kind of good they take reflective equilibrium to achieve. On one interpretation reflective equilibrium merely indicates that a reasoner has done what is rationally required of them relative to their initial data (e.g., a set of intuitions about certain logical inferences), but it would take further argument to show that a reasoner's doxastic attitudes are also likely to be true. Under this interpretation reflective equilibrium is a rational ideal regarding the *internal* coherence of doxastic states rather than truth-conducive rationality.[24]

4 The Dilemma of Logical Akrasia

The upshot of §3 is what we might call the *Dilemma of Logical Akrasia*:

> Insofar as we take our logical theories to be appropriately formal and complete, then either:[25]

(i) we must be agnostic about the consistency of PA (on the pain of triviality), which would be extremely odd at best;

(ii) or we must accept being logically akratic, i.e., accept that we are trapped in an inescapable, infinite hierarchy of logical akrasia.

Notice that while taking the second horn of the dilemma doesn't rule out the existence of a rational fixpoint somewhere on the theoretical ladder, it does eliminate the possibility of an akrasia-free state which is accessible to us (since Gödel's incompleteness results range over all axiomatizable theories).

[24] For further discussion of positive epistemic evaluations and their connection to truth-conduciveness, see e.g., (BonJour, 1985; Alston, 1989; Littlejohn, 2012; Berker, 2013).

[25] In this context the term '*complete*' should be understood as follows: *Within a given domain, every question is answerable, i.e., for any φ in the domain, it holds that φ or not-φ* . This kind of completeness is also known as 'Syntactic Completeness'. Note that Syntactic Completeness doesn't entail that every particular answer has got the same epistemic status. The point is merely to suggest that we are committed to completeness in the sense that the question of whether PA is consistent has got an answer, but this is definitely not committing us to a logical theory which can decide every question.

Different interpretations of the term '*formal*' in the context of logical theorizing can be found in (MacFarlane, 2000; Dutilh Novaes, 2012).

Further, the Dilemma of Logical Akrasia is special in at least two ways. First, it involves an *unsolvable* case of logical akrasia while most cases of logical akrasia are clearly solvable, e.g., by converting to a fully classical theory. If the intuitionist we met in §2 were willing to convert to a fully classical theory, then their akratic state would dissolve. But the case of PA is different as it looks more similar to an *epistemic blindspot*; where proposition $\langle p \rangle$ is an epistemic blindspot for subject S at time t if and only if $\langle p \rangle$ is consistent but unknowable by S at t (Sorensen, 1988, 2020). Similarly, the consistency presumption is fundamental to our use of PA, but it just cannot be proved (and thus known) from within the bounds of the theory itself.

Second, unlike the akratic issues we focused on above (cf. §2), the Dilemma of Logical Akrasia cuts across the divide between classical and non-classical logicians. In the case of PA it seems that we are all either agnostic (on the pain of triviality) or akratic!

So, in the end, taking the first horn doesn't sit well with our general intellectual outlook because we want to avoid being agnostic about the consistency of PA; on the other hand, going for the second horn is an unpleasant move as it reveals a boundary on logical theorizing which seems to conflict with our rational ideals (e.g., reflective equilibrium).

What exact consequences the dilemma has for epistemic rationality in general, we'll leave for future research to unravel.

5 Epilogue: Escaping the Dilemma of Logical Akrasia

In this short epilogue we'll consider a quick and dirty proposal of how one can translate the Dilemma of Logical Akrasia into a case of regular epistemic akrasia; and further how one might escape the dilemma when it's spelled out this way.

Let's first reformulate the Dilemma of Logical Akrasia in terms of *beliefs* such that it becomes a case of epistemic akrasia. Spelled out in terms of premises and conclusion(s), we get:

1. S believes PA [by Indispensability].
2. S believes Con_{PA} [by No-Miracles].
3. S believes Con_{PA} is a logical principle [in absence of reasons to the contrary].
4. S believes $\nvdash_{PA} Con_{PA}$ [by Incompleteness].
5. S believes in the strong interpretation of reflective equilibrium with respect to PA: It's permissible for S to believe a logical principle only if PA proves it.
6. Therefore: S believes Con_{PA} and believes that $\langle$it's forbidden to believe $Con_{PA}\rangle$.

7. Ergo: S is epistemically akratic.

Now, it seems fair to suggest that we don't want to consider rejecting premises (1), (2), and (4); which leaves us with the possibility of rejecting one or both of (3) and (5) in order to escape the dilemma. That is to say:

1. ~~S believes PA [by Indispensability].~~
2. ~~S believes Con_{PA} [by No-Miracles].~~
3. S believes Con_{PA} is a logical principle [in absence of reasons to the contrary].
4. ~~S believes $\nvdash_{PA} Con_{PA}$ [by Incompleteness].~~
5. S believes in the strong interpretation of reflective equilibrium with respect to PA: It's permissible for S to believe a logical principle only if PA proves it.
6. Therefore: S believes Con_{PA} and believes that $\langle$it's forbidden to believe $Con_{PA}\rangle$.
7. Ergo: S is epistemically akratic.

Consider first the possibility of rejecting (3), i.e., the logicality of Con_{PA}. The principle stated by Con_{PA} is certainly a well-formed sentence of PA; and in that specific sense it is logical. It also expresses something essential to PA (or at least our use of PA). But perhaps Con_{PA} is still not logical in the right way for strong reflective equilibrium to apply to it. Is it really fair to expect strong reflective equilibrium to apply to paraconsistent logical theories, for example?

Consider next the possibility of rejecting (5) instead. While reflective equilibrium may be initially plausible when viewed as a philosophical *method* or as "the prima facie epistemology of logic" (Cohnitz and Estrada-González, 2019, p. 137), it does seem like an overly demanding *output* to expect from applying the method in the context of logical theorizing and PA. Gödel's theorem already suggests that epistemic principles of that kind are hopeless, because every logical theory with the same expressive power as PA (or more) has a Gödel sentence. Why insist on something impossible? (This is an "ought-implies-can" violation perhaps). Moreover, what reason do we have to think that strong reflective equilibrium is a norm of belief within logical theorizing? It seems that you'll need overly demanding bridge principles to establish the right connections between logic and epistemology in order to get this going (cf. (MacFarlane, 2004)). Another way to put this point: the dilemma relies on it being an epistemic ideal that there is a reflective equilibrium between what one is committed to and what one's accepted theory can prove (call it 'RE'). An alternative, and perhaps more plausible ideal is that there be a reflective equilibrium between what one's committed to and what one's epistemic practice can justify (call it 'RE*'). What speaks in favor of RE over RE*?

Supposing that at least one of the rough strategies outlined above is successful in letting us escape the dilemma—when it's framed in terms of epistemic akrasia—we are thus left to ask whether it's still a problem if S is *logically* akratic after rejecting either of these premises and avoiding epistemic akrasia. Some logicians of the post-Gödel era may simply shrug their shoulders and bite the bullet here. In a way what Gödel's second incompleteness result tells us is that we can't both have consistency and syntactic completeness when it comes to theories with a certain amount of expressive power. So, perhaps logical akrasia is simply something that working logicians have come to live with in the aftermath of Gödel. They may also want to suggest that there is an important difference between the cases of logical akrasia exemplified by the intuitionist rejecting the excluded middle (cf. §2) and the specific case of PA. In the former, the intuitionist can't combine their official theory and the background logic they are committing to into a jointwise consistent whole, whereas this is certainly possible in the latter—it's just that the background logic must be stronger than the theory PA itself.

Chapter 7

The Epistemology of Disagreement Revisited

We have now reached the seventh chapter of the monograph. In this chapter we'll look back at some of the ground we have covered thus far—viz., the initial literature review of the epistemology of *peer disagreement* (cf. Introduction) and the paradigmatic case of *deep disagreement*, **Young Earth Creationist** (cf. Chapter 1)—and draw some conclusions about these themes in light of what we learned from our three interpretations of logical disagreement from previous chapters. In general outline the chapter is split into two main sections. In §1 we'll revisit the epistemology of peer disagreement and argue that the epistemic significance of central principles from the literature are at best deflated when applied in the context of logical disagreement. The cumulative outcome of section §1 is thus a skeptical pressure against sweeping answers to the Doxastic Disagreement Question: *What is the epistemically rational response to cases where one disagrees with an epistemic peer as to whether* $\langle p \rangle$*?* Since it is not even possible to give a normatively satisfying answer in the cases where $\langle p \rangle$ happens to be a *logical* proposition, we can't give a completely general answer either. On the heels of this, §2 develops a simple formal model of paradigmatic deep disagreement in a refined Horty-style default logic and compares the result with some obvious competitors. We'll see that our simple model fares quite well in comparison to both Classical Propositional Logic and Subjective Bayesianism. Finally, we'll relate our discussions of various formal models of deep disagreement to our general discussion of logical disagreement. We conclude that *if* logical disagreements are indeed structurally similar to deep disagreements, then how we ought to respond to them, epistemically speaking, will depend on whether it is legitimate to have an entirely subjective ranking of logical principles.

Keywords Epistemology of Disagreement; Logical Disagreement; Epistemic Peerhood; Rational Uniqueness; The Independence Principle; Model-Building in Philosophy; Deep Dis-

agreement; Subjective/Objective Rankings

1 The Epistemology of Peer Disagreement Revisited

1.1 Peerhood Revisited

In our introduction to the epistemology of disagreement we met the central idealization *Epistemic Peerhood* for the first time, and we saw how standard views from the literature are restricted to cases where epistemic peerhood is in place. Recall that according to Kelly's influential definition of peerhood, two agents are epistemic peers exactly when:

1. they are equals with respect to their familiarity with the evidence and arguments which bear on that question, and;

2. they are equals with respect to general epistemic virtues such as intelligence, thoughtfulness, and freedom from bias.[1]

As we know, the peerhood-model of disagreement is meant to establish a certain kind of epistemic symmetry between the interlocutors. The model aims to remove the confounding possibility of it being epistemic asymmetries between the disagreeing parties—rather than a supposed significance of the disagreement itself—driving our assessments of central cases.

One clear reason to criticize the peerhood-model is that disagreements with non-peers *can* be epistemically significant in some cases. Suppose that I am a medical doctor with a known track record of getting a certain type of diagnosis right 80% of the time, while my junior colleague has a known track record of 60% right answers with respect to the same type of diagnosis. It then turns out that we disagree about a particular patient; I am initially quite confident that the patient has the diagnosis in question, but my colleague is sure that this is not the case. Intuitively, this should make me reduce my confidence that I am right—after all, my colleague is right 60% of the time, and this should give me some reason to think that I might have made a mistake although my junior colleague is not my peer on this question (Kappel and Andersen, 2019, p. 1107).

Another reason to criticize the peerhood-model—which is much more pivotal to us—is that while peerhood may be a fair and useful idealization in standard cases such as **Restaurant**

[1] Kelly (2005) notes the familiar fact that, outside of a purely mathematical context, the standards of equality between two entities, along some dimension, are highly context-sensitive. Thus, whether two individuals count as epistemic peers will depend on the specific standards for epistemic peerhood within a given context. In the same way, whether two individuals count as 'the same height' will depend on the specific standards of measurement that are in play, see, e.g., (Lewis, 1979).

and **Horse Race**, it doesn't have the intended epistemic effect in the context of logical disagreement. To see this, let's consider a case where logicians Graham Priest and Tim Williamson are disagreeing over the validity of Modus Ponens, i.e., the proposition $\langle \varphi \rightarrow \psi, \varphi \vDash \psi \rangle$. Williamson believes that Modus Ponens is valid (because of his commitment to classical logic, call it 'CL'), while Priest doesn't (as he endorses the Logic of Paradox, 'LP'). Now, it's surely not crazy to suggest that Williamson and Priest are equally familiar with the relevant evidence and arguments that bear on the validity of Modus Ponens, and further that they are equally virtuous in terms of intelligence, thoughtfulness, and freedom from bias; and if this much is true, then they live up to (1) and (2) from Kelly's definition. Hence they will count as each other's epistemic peers. Nonetheless, it is quite difficult to see how the fact of the disagreement by itself provides a (good) reason for either side to conciliate—as would allegedly be the case in standard scenarios of peer disagreement like **Horse Race**. The higher-order evidence generated by the fact that Tim and Graham disagree as to whether Modus Ponens is valid doesn't seem to act as a defeater to their initial doxastic attitudes regarding the target-proposition—viz., $\langle \varphi \rightarrow \psi, \varphi \vDash \psi \rangle$—because in this case each side of the dispute has excellent, well-developed explanations of why they disagree, and why it is thus rational for them to stick to their guns.

As we saw earlier (cf. Introduction), much attention in the peer disagreement debate has been paid to *symmetry breakers*, i.e., facts that will allow at least one party of peer disagreement to rationally assign more weight to their own epistemic position. Epistemologists with steadfast leanings have tried to come up with plausible ways to break the epistemic symmetry between the peers involved in disagreements. Yet, as we can see from the Modus Ponens-example above, the relevant kind of epistemic symmetry is actually lacking in some very realistic cases of logical disagreement, involving well-known logical theories, and this is even true when the involved parties are highly competent and each other's peers.

And even if we acknowledge that according to state of the art research on higher-order evidence, the label 'higher-order evidence' has been used equivocally to refer to (a) evidence about the rationality of one's belief; (b) evidence about one's reliability; (c) evidence about what evidence one has; and (d) evidence about what one's evidence supports (Ye, 2023, p. 3). It is still very hard to see how this would change the steadfast outcome of the dispute between Priest and Williamson. Consider in turn the readings (a)-(d) of 'higher-order evidence' applied to the Modus Ponens-example.

If we opt for reading (a), the disagreement between Priest and Williamson generates evidence about the rationality-status of their initial doxastic attitudes *vis-à-vis* $\langle \varphi \rightarrow \psi, \varphi \vDash \psi \rangle$. Yet, as both Priest and Williamson individuate logical evidence holistically (cf. Chapters 1 & 4), their initial attitudes toward $\langle \varphi \rightarrow \psi, \varphi \vDash \psi \rangle$ will be rational given their respective choice of logic or logical theory. There is, for instance, nothing surprising or irrational about Williamson's believing that Modus Ponens is valid relative to CL; and that Priest disagrees with this is readily explained by his commitment to LP. Of course, there is the worry lurking in the background that one side might be right about their logical theory capturing the "One True Logic", while the other side of the dispute is misguided in preferring a logic that differs from the ultimately correct one. This notwithstanding, when we are just considering Priest and Williamson's strife about the validity of Modus Ponens, and we (correctly) assume that they individuate logical evidence holistically, then steadfastness on either side looks straightforward.

Suppose we go for reading (b) of 'higher-order evidence' instead. Then the dispute between Priest and Williamson would provide evidence about their respective reliability-profiles *vis-à-vis* assessments of deductive validity rather than the rationality-status of their initial doxastic attitudes toward $\langle \varphi \rightarrow \psi, \varphi \vDash \psi \rangle$. In this light, we can see that the case of logical disagreement between our two epistemic peers isn't at all like a case, where someone learns they have ingested a reasoning-distorting drug or are suffering from hypoxia (Christensen, 2010). That is, a case where one cannot trust oneself to be reliable with respect to a given (class of) question(s). Rather the tension between the doxastic attitudes of the disagreeing parties is exactly the expected outcome of reasoning from the two different logics (or logical theories), viz., CL and LP, to the verdict whether Modus Ponens is valid or not. Both sides of the dispute seem as (un)reliable as before they learned about their concrete disagreement with each other.

Next, take reading (c) of 'higher-order evidence'. According to this interpretation, the logical dispute between Williamson and Priest is evidence about what evidence each party has in their possession. Again—unsurprisingly—the outcome will be that Williamson possesses first-order evidence that is relativized to CL, while Priest possesses first-order evidence which is relative to LP. This is simply what follows from their holistic individuations of logical evidence. So, again, the presence of the relevant dispute doesn't provide either side with a (good) reason to revise their initial doxastic attitude with respect to the target-proposition concerning the validity of Modus Ponens.

Consider then finally reading (d). On this interpretation, the fact of disagreement between Williamson and Priest is evidence about what their evidence supports. As the reader will have guessed by now, there are no big surprises following from this interpretation of higher-order evidence either. The case of disagreement between Priest and Williamson is completely transparent when it comes to their logical evidence for and against Modus Ponens, and it is clear to both sides that their evidence is relativized to LP and CL, respectively. So—keeping the holistic individuation of logical evidence fixed—we'll have the steadfast outcome once again.

What we can conclude based on the preceding paragraphs of this subsection is ironically enough that while many authors in the peer disagreement debate have been on the look for *symmetry breakers* in response to standard cases such as **Horse Race** and **Restaurant**, what is needed for some very simple cases of logical disagreement is really *symmetry makers*! In order to invoke epistemic symmetry in logical disputes like the Modus Ponens-case just considered we need a non-partisan notion of validity, i.e., a notion that is not internal (or relativized) to any particular theory of logic, but as we know from previous chapters, the current default position is that no such notion is available. Contemporary philosophers of logic are, by default, *holists* about logical evidence and the epistemic justification of logical propositions. While this doesn't necessarily show that we should all become *atomists* regarding these matters, it does suggest that the peerhood-model of disagreement won't have the intended epistemic effect on disagreements in or about logic unless we make special assumptions that run counter to the current standards of philosophy of logic.

1.2 Uniqueness Revisited

Reconsider now Rational Uniqueness, which was thoroughly discussed in Chapter 2 of the monograph, and which is also known as the Uniqueness Thesis ('UT'):

> For any body of evidence E and proposition $[p]$, E justifies at most one doxastic attitude toward $[p]$. (Matheson, 2011, p. 360)

As we have learned, UT features frequently in debates concerning the possibility of rational peer disagreement: *If two epistemic peers disagree as to whether $\langle p \rangle$, is it then possible for both of them to be propositionally justified in their incompatible doxastic attitudes toward $\langle p \rangle$?* If UT is true, the answer is negative.

In spite of this established background, it's actually quite hard to go from standard analyses of peer disagreement cases like **Restaurant** and **Horse Race** to analyses of *logical* disagreement in particular; and it turns out to be rather unclear how much "heavy lifting" UT can do when it comes to analyzing even very simple cases of logical disagreement, such as the Modus Ponens-case from §1.1 above. For our purposes it is, again, particularly interesting to observe that if one's *logical evidence* is relative to one's preferred logic or logical theory—which, as we said, is the current default position—then the epistemic significance of UT is seriously deflated.

As underscored in Chapter 2, UT is supposed to motivate a certain response to peer disagreement, namely that at most one peer can be propositionally justified in such situations. But if logical evidence is relativized to the logical theory of one's choice, the scope of UT is reduced drastically. You can now only share logical evidence with those from your own theoretical equivalence class, and there can be as many of those classes as there are acceptable logical theories. Yet, this is in no way what the UT-proponent wanted from their thesis in the context of the epistemology of disagreement. For let us remind ourselves of how strong a thesis UT is supposed to be: it concerns all bodies of evidence, no matter what subject possesses it, and no matter the time.

Further—considering the popularity of modern versions of Logical Pluralism (Beall and Restall, 2000, 2006; Russell, 2019b)—it is tempting to relate the issues surrounding logical evidence and uniqueness to some more established debates about permissible *epistemic standards* (Titelbaum and Kopec, 2019). Plenty of formal epistemologists claim that a body of evidence supports a hypothesis only relative to a rational reasoning method, and since there are multiple, extensionally non-equivalent rational reasoning methods available, there is not always an unambiguous fact of the matter about whether some evidence supports a particular hypothesis. Subjective Bayesianism, for example, could deny UT by appeal to legitimate differences in epistemic standards. In general, Bayesians hold that any rational agent's credences at a given time can be obtained by conditionalizing their *hypothetical prior* ('Cr_h') on their total evidence at that time. For a total body of evidence E and a hypothesis H, the evidence supports the hypothesis exactly when $Cr_h(H \mid E) > Cr_h(H)$. Here, facts about evidential support are therefore relative to the hypothetical prior of the agent in question, and we can plausibly think of an agent's hypothetical prior as capturing their epistemic standards. Some Objective Bayesians claim that there is a unique rational hypothetical prior, so, in their case—while evidential support is relative to the hypothetical prior—there is still at most one rational hypothetical prior, and so UT is true. Yet some Subjective Bayesians believe that multiple hypothetical priors are rationally acceptable. Thus, for them, two rational agents could have different hypothetical priors—representing different epistemic standards—and we could have situations where the same body of evidence E supports a hypothesis H for one of them while it doesn't for the other. Hence, for some bodies of evidence and some hypotheses, there are no justificational facts of the sort UT asserts, i.e., there are no facts about simple, non-relativized, support-relations as UT would have it.

So, to sum up the findings of this subsection, while UT is thought to have a central role to play in mainstream epistemology of disagreement, the thesis comes under a serious pressure when applied to even simple cases of logical disagreement. If logical evidence is taken to be relative to preferred logical theory, and support-relations can vary with legitimate differences in epistemic standards, then the epistemic significance of UT is deflated.

1.3 Independence Revisited

Next, let's consider whether the widely discussed Independence Principle is applicable and useful in the context of logical disagreement. As we have seen previously (cf. Introduction), the original version of the principle says:

> *The Independence Principle*. In evaluating the epistemic credentials of another's expressed belief about $\langle p \rangle$, in order to determine how (or whether) to modify my own belief about $\langle p \rangle$, I should do so in a way that doesn't rely on the reasoning behind my initial belief that $\langle p \rangle$ (Christensen, 2009) (slightly altered

notation).[2]

But suppose now that a classical logician believes that $\langle \neg\neg\varphi \vDash \varphi \rangle$, while an intuitionistic logician denies this proposition. As should be clear from the context, the truth of $\langle \neg\neg\varphi \vDash \varphi \rangle$ is logically dependent on the excluded middle, i.e., $\langle \vDash \varphi \vee \neg\varphi \rangle$, which is *exactly* what the classicist and the intuitionist are disagreeing about to begin with!

Hence, in one of the most famous cases of logical disagreement, the Independence Principle seems inapplicable (and perhaps even irrational).

Another example—to strengthen the case against the Independence Principle in the context of logical disagreement—comes from proof-theory rather than semantics. Suppose an intuitionistic proof-theorist believes $\langle \varphi \vee \psi, \neg\psi \vdash \varphi \rangle$ while a proponent of Minimal Logic rejects this. In this example, the target-proposition under dispute $\langle \varphi \vee \psi, \neg\psi \vdash \varphi \rangle$ is logically dependent on Ex Falso Quodlibet, i.e., $\langle \varphi \wedge \neg\varphi \vdash \psi \rangle$, which is exactly the difference between the two sides of the dispute to begin with! (Restall and Standefer, 2023, p. 71).

What these cases show is that independence in the above sense isn't something one can expect to find in logical disagreements. Rather one should expect *theory-dependence* as there is no easy way for the parties involved in logical disagreements like those we've just considered to avoid reasoning from the basic theoretical commitments that led them to their initial doxastic attitude towards $\langle p \rangle$. So—a central lesson reappears once more—it seems what we really need in some cases of logical disagreement is *symmetry makers* rather than *symmetry breakers*.

In the main argument of Chapter 4 we saw that one potential symmetry maker between the sides of logical disagreements is given by the truth of the E-literal about a liberal version of Universal Instantiation, i.e., $[\forall x Px, \Gamma \vDash Pa]$. Since $[\forall x Px, \Gamma \vDash Pa]$ is a foundational E-literal of deduction—semantically understood—which must be presupposed by any plausible logical theory, it might act as a symmetry maker in certain logical disputes.

On the other hand, if we also relate our discussion of the Independence Principle to *intra-personal* logical disputes (cf. Chapter 6), it's interesting to observe how states of logical akrasia might act as symmetry-breakers in some cases of *inter-personal* logical disagreement. If, for example, Williamson knows that Priest's commitment to Modus Ponens in his background logic makes his official acceptance of LP logically akratic, then we might think of this internal incoherence as an epistemic symmetry-breaker in Williamson's favor in the Modus Ponens-case. Yet, as the reader will remember from our general investigation of akrasia in chapters 5 and 6, it's not entirely straightforward to say how devastating a charge with logical akrasia really is.

[2] The Independence Principle—as stated above—is vague in multiple ways; see (Christensen, 2019) for Christensen's latest revisions of and thoughts about the principle.

1.4 Conclusion of §1

The cumulative outcome of this section is a skeptical pressure against sweeping answers to the Doxastic Disagreement Question: *What is the epistemically rational response to cases where one disagrees with an epistemic peer as to whether $\langle p \rangle$?* Since it is not even possible to give a normatively satisfying answer in the cases where $\langle p \rangle$ happens to be a logical proposition, we can't give a completely general answer either.

Thus, we end up agreeing with the skeptical attitude of Srinivasan and Hawthorne, which we met in the very beginning of the monograph:

> We have suggested that those of us who hope for a general and intuitively satisfying answer to the question that is at the centre of the disagreement debate—namely, what we ought to do, epistemically speaking, when faced with disagreement—might be hoping in vain. There are deep structural reasons why such an answer has proven, and will continue to prove, elusive. (2013, p. 28)

As we have just seen in §§1.1-1.3, a number of central principles from the peer disagreement-literature that are all thought to be epistemically significant, and of great import in answering the Doxastic Disagreement Question, are at best deflated in the context of *logical* disagreement. So, alas, the negative conclusion to draw here is that any hope of finding a sweeping and normatively satisfying answer to the Doxastic Disagreement Question is in vain. For even some very simple examples of logical disagreement have made this much clear.

2 The Epistemology of Deep Disagreement Revisited

Now, as it has been suggested at several points of the monograph already, logical disagreements are often importantly different from mundane cases of disagreement like **Restaurant** and **Horse Race**. Oftentimes logical disagreements are more similar to theory-loaded quarrels from the philosophy of science, e.g., between Darwinians and Creationists, proponents of Skinner's Behaviorist theory of language and Chomsky's Cognitivism etc. As we said, logical disagreements may in fact bear a close structural resemblance to *deep disagreements*, where we disagree about our fundamental epistemic principles (cf. Chapter 1). Using an analogy for illustration, we could put the point as follows: when in logical disagreement it's commonly not the case that we are simply disagreeing over a surface-level proposition, such as the fair share of our restaurant bill, rather our dispute revolves around the completely general mathematical principles underlying each of our individual calculations of splitting the bill fairly. Or, to put the same point differently, when in logical disagreement we are frequently not just disagreeing about whether horse A or B crossed the finish line first, but rather the fundamental reliability of visual perception as a belief-forming process in general.

Below we'll take a first stab at building a formal model of deep disagreement in order to assess the epistemological consequences of the proposed structural similarities between logical and deep disagreement more precisely. We'll find inspiration in artificial intelligence and use the formal machinery of *default logic*. Default logic has been a very active research topic in artificial intelligence ever since the early 1980s, but has not received as much attention in the philosophical literature thus far. This section shows one way in which the technical tools of artificial intelligence can be applied in contemporary epistemology by modeling a paradigmatic case of deep disagreement using default logic. In §2.1 model-building viewed as a kind of philosophical progress is briefly motivated, while §2.2 (re)introduces the case of deep disagreement we aim to model, viz., the **Young Earth Creationist**. Following this, §2.3 defines our formal framework: a refined Horty-style default logic. §2.4 then uses the defined framework to model deep disagreement, while §2.5 provides a critical discussion of the result. Finally, §2.6 relates the discussion of our formal model of deep disagreement to the overarching theme of logical disagreement. We conclude that *if* logical disagreements are indeed structurally similar to deep disagreements, then how we ought to respond to them, epistemically speaking, will depend on whether it is legitimate to have an entirely subjective ranking of logical principles.

2.1 Model-Building in Philosophy

It is uncontroversial that model-building is crucial to the progress of science. When a certain phenomenon cannot be studied directly, for whatever reason, building a (formal) model of it can often lead to progress indirectly.[3] Studying a model in detail may give rise to new insights about the modeled phenomenon, and these insights can eventually result in a better model than the one we started out with.

Curiously, however, the gradual process of model-building is perhaps not as celebrated in philosophy as in science—even though building better and better models of complex phenomena is an integral part of philosophical progress as well (Williamson, 2007, 2013a, 2017a, 2019). As Williamson observes:

> [I]n philosophy, too, one form of progress is the development of better and better models... The need for model-building is hardest to avoid where the complex, messy nature of the subject matter tends to preclude informative exceptionless universal generalizations. The paradigm of such complexity and mess is the human world. Hence the obvious places to look for model-building in philosophy are those branches most distinctively concerned with human phenomena, such as ethics, epistemology, and philosophy of language. (Williamson,

[3] Astrophysics being an illustrative example.

> 2017a, pp. 160-163)[4]

Social epistemology fits the bill here. The complex, multi-agent dynamics found in core topics of the field such as group rationality, expert testimony, peer disagreement, epistemic injustice etc., naturally lend themselves to systematic and intuition-guiding models.

In what follows we aim to take a first stab at formally modeling *deep disagreement*, and to this end we'll use the formal machinery of *default logic*. A reason in favor of using this framework for modeling exercises in social epistemology is the common interpretation of default rules (or simply defaults) as *defeasible generalizations* (Horty, 2012)—i.e., exactly the kind of generalizations Williamson takes to be prevalent in those branches of philosophy most distinctively concerned with human phenomena. An example is: *If Tweety is a bird, then Tweety can fly*. Clearly, learning the truth of the antecedent provides a reason to believe the consequent, but additionally learning that Tweety is a penguin defeats it.[5]

Before getting down to business, it's worth stressing that our aim is rather modest. Our ambition is merely to construct a provisional model of paradigmatic deep disagreement, which is open to—perhaps even in need of—further innovation. Yet, our modesty should not be confused with a lack of ambition as it encapsulates the spirit of model-building very well. Just like in science; model-building in philosophy is an incremental achievement.

2.2 Deep Disagreement Revisited

Let's now get an intuitive grasp of deep disagreement, which is the phenomenon that we want to model. Consider once again the **Young Earth Creationist** (cf. Chapter 1):

> Henry is an Evangelical young Earth creationist, who accepts that the Earth is no more than 6000 years old and a nexus of conspiratorial claims as evidence of why scientists have been misleading us about the age of the Earth. Henry also rejects the theory of evolution and contemporary cosmology, citing literal readings of the Bible: 'your denial of scripture is unjustified', he says. Henry's neighbor Richard is a proponent of so-called 'New Atheism', and rejects the religious and young Earth creationist views of his neighbor Henry, and asserts

[4] Of course Williamson's metaphilosophy isn't completely uncontroversial. Consider the case of philosophy of language, for example: if model-building in this area means understanding natural language and linguistic phenomena via semantic models, then many philosophers (and linguists) would disagree with Williamson's metaphilosophical view.

[5] This defeasibility makes default logic a *non-monotonic* framework. Monotonic logics—such as classical logic—satisfy the property that for any well-formed formula φ from the language L, if $\varphi \in L$ is a consequence of a set of formulas $\Gamma \subseteq L$ and if $\Gamma \subseteq \Delta \subseteq L$, then φ is also a consequence of Δ. Non-monotonic logics, by contrast, allow conclusions to be withdrawn in the light of new information.

> that the Earth is much older than 6000 years: 'your denial of geology and evolutionary biology are unjustified', he says. (Ranalli, 2021, p. 984)

As mentioned earlier, this case has been widely discussed in the literature and is considered a paradigmatic case of deep disagreement (Lynch, 2010; Pritchard, 2010; Kappel, 2012; Hazlett, 2014; Ranalli, 2021; Ranalli and Lagewaard, 2022a,b).

Although there are several different ways of understanding the essentials of deep disagreement, we'll focus on the so-called *Fundamental Epistemic Principle Theory* to avoid unnecessary detours.[6] According to this theoretical stance, deep disagreements are *deep* because they aren't solely concerned with "surface-level" propositions about, say, a particular weather forecast (Christensen, 2007), but also propositions stating the fundamental epistemic principles we ought to apply when trying to predict the weather in general. In other words, deep disagreements are disagreements over fundamental epistemic principles like those specifying which traditions, institutions, methods, sources of evidence, and patterns of reasoning to rely upon when forming beliefs (Kappel and Andersen, 2019).

Rational irresolvability is often considered a necessary property of deep disagreements because of their dialectical setup (Wittgenstein, 1969b; Fogelin, 2005; Lynch, 2010, 2016; Kappel, 2012). How is one supposed to give a compelling argument for target-proposition $\langle p \rangle$, when one's interlocutor asserts $\langle \text{not-}p \rangle$ (or suspends judgement as to whether $\langle p \rangle$), and does so by appealing to fundamental epistemic principles that conflict with one's own?[7] In the words of Michael Lynch:

> ...explicit defenses of such principles will always be subject to a charge of circularity. Hume showed that the principle of induction is like this: you can't show that induction is reliable without employing induction. It also seems true of observation or sense perception. It seems difficult, to say the least, to prove that any of the senses are reliable without at some point employing one of the senses. Similarly with the basic principles of deductive logic: I can't prove basic logical principles without relying on them. In each case, I seem to have hit rock bottom... (Lynch, 2016, pp. 250-251)

As should be clear—in the case **Young Earth Creationist**—Henry and Richard disagree about the age of the Earth at surface-level, but their disagreement depends on a much more

[6] According to Ranalli (2021), state of the art research on how to best characterize deep disagreement falls roughly into two theoretical camps. On the one hand we have the *Hinge Proposition Theorists* (Wittgenstein, 1969b; Feldman, 2005a; Fogelin, 2005; Friemann, 2005; Hazlett, 2014); on the other the *Fundamental Epistemic Principle Theorists* (Lynch, 2010; Kappel, 2012; Jønch-Clausen and Kappel, 2015; Lynch, 2016; Kappel, 2021; Lagewaard, 2021).

[7] See (Ranalli, 2020) for a helpful disambiguation of the term 'rationally irresolvable'. Consult (Martin, 2021c) for a recent argument *against* the rational irresolvability of deep disagreements.

fundamental disagreement about evidential standards and what justifies beliefs. This is why their story has come to be viewed as a paradigmatic case of *deep* disagreement.

2.3 Default Logic

Default logic has been a very active research topic in artificial intelligence since the early 1980s (Reiter, 1980; McDermot and Doyle, 1980; Reiter and Criscuolo, 1981; McCarthy, 1986; Poole, 1988; Brewka, 1989; Baader and Hollunder, 1993; Brewka, 1994a,b; Makinson, 1994; Baader and Hollunder, 1995; Rintanen, 1995; Antoniou et al., 1996; Rintanen, 1998; Antoniou, 1999; Brewka and Eiter, 2000; Antonelli, 2005; Thomason, 2018), but has not received as much attention in the philosophical literature thus far.[8]

Nonetheless, John F. Horty's monograph *Reasons as Defaults* (2012) highlights several promising applications of default logic in philosophy—e.g., modeling the structure and strength of reasons, defeaters, and arguments.[9] This subsection refines the basic definitions of Horty's default logic such that it can model the multi-agent dynamics of typical deep disagreement scenarios.

Horty's Framework

In its most basic form Horty's default logic is simply classical propositional logic extended with default rules.

Let Φ be a countable set of atomic propositions and $\mathcal{L}$ a language such that:

$$\varphi := p \mid \top \mid \neg\psi \mid \psi \rightarrow \psi'$$

When $\Gamma \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}$, we write $\Gamma \vdash \varphi$ to express left-to-right classical deducibility. Denote the logical closure of Γ by $Th(\Gamma) := \{\varphi : \Gamma \vdash \varphi\}$. Where $\varphi, \psi \in \mathcal{L}$, a default rule is any expression of the form:

$$(\varphi \rightsquigarrow \psi)$$

It's important to notice that default rules are metalinguistic, so they cannot be expressed in $\mathcal{L}$. Further, the symbol '$\rightsquigarrow$' cannot be nested to generate more complex default rules.

[8] Yet, it's worth flagging that philosophical works involving default logic has become more common in recent years, see, e.g., (Bonevac, 2018; Knoks, 2021a,b, 2022).

[9] See also (Horty, 2007a,b, 2016).

We let $\mathcal{D}$ denote the set of all possible defaults (with typical elements δ, δ'...). For a default rule $\delta = (\varphi \rightsquigarrow \psi)$, let $Conclusion(\delta) := \psi$. And for a set of default rules $D \subseteq \mathcal{D}$, let $Conclusions(D) := \{Conclusion(\delta) : \delta \in D\}$.

Consider next a rational agent's basis for default reasoning. A single agent default theory is a tuple:

$$\Delta = (W, D, \leq, \Gamma)$$

W denotes the agent's set of background information, i.e., hard facts; D refers to the set of default rules which are *available* to the agent (these need not be plausible defaults, just available ones). The order $\leq$ is a non-strict partial order on D with the formal properties transitivity, reflexivity, and antisymmetry.[10] Suggesting the following reading of $\delta \leq \delta'$: "δ' *represents a default of a priority which is at least as high as the one* δ *represents*," where "*higher priority*" means less easily defeasible. We say that $\delta \in D$ is *fundamental* when there is no $\delta' \in D$ such that $\delta < \delta'$, i.e., when there is no other available default δ' of strictly higher priority than δ in D.[11]

A **scenario** S (based on a default theory Δ) is a subset $S \subseteq D \subseteq \mathcal{D}$ contained in Δ. We interpret S as a particularly plausible set of available default rules, i.e., the defaults of which the antecedents provide sufficient support for their conclusions according to the agent in question. We'll assume that if a given agent considers δ fundamental, then $\delta \in S$ also holds for that agent.

The last element of the tuple (i.e., the default theory) is the agent's belief set Γ. Define a belief set:

$$\Gamma = Th(W \cup Conclusions(S))$$

That is, the logical closure of the hard background information plus the conclusion-set of the plausible defaults available to the agent.

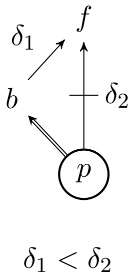
To illustrate, consider default theory Δ such that $W = \{p\}, D = \{\delta\}$ with $\delta = (p \rightsquigarrow \neg q)$, and $\leq = \emptyset$. Assuming that δ is plausible, the resulting belief set is $\Gamma = Th(\{p, \neg q\})$.[12]

[10] Relation R is transitive if and only if $\forall x \forall y \forall z((Rxy \wedge Ryz) \rightarrow Rxz)$. R is reflexive if and only if $\forall x Rxx$. R is antisymmetric if and only if $\forall x \forall y((Rxy \wedge Ryx) \rightarrow x = y)$.

[11] Notice that this understanding of fundamentality allows a reasoner to have multiple fundamental defaults as long as they are of equal priority.

[12] Horty *doesn't* directly associate extensions of default theories with beliefs (Horty, 2012, pp. 34-40): a default theory Δ may have no extensions or multiple ones, and identifying the Δ-beliefs with *the* extension of Δ is therefore not well-defined. Horty discusses both multiple and empty extensions, but he does not give a clear solution. As we won't be confronted with empty extensions in this chapter, we simply ignore that problem. For

Another—slightly more sophisticated—example is a classic of non-monotonic reasoning. The example concerns the bird Tweety and its ability to fly. The fact that Tweety is a bird provides a reason to conclude that Tweety can fly. But if Tweety is also a penguin, the reason to conclude that Tweety can fly is defeated. The details of the Tweety-example is captured in Figure 7.1 below.



Propositions:

b : Tweety is a bird. f : Tweety flies.
p : Tweety is a penguin.

All possible scenarios:

$\emptyset$,
$S_1 = \{\delta_1\}$, $S_2 = \{\delta_2\}$,
$S_3 = \{\delta_1, \delta_2\}$

Figure 7.1: The Tweety Triangle. Circled propositions constitute the set of hard background information; the double arrow shows that $(p \rightarrow b)$ is in the background information. A δ-labeled arrow from one formula φ to another ψ means the default $\delta = (\varphi \rightsquigarrow \psi)$ is among the available defaults. When a δ-labeled arrow from φ to ψ is crossed out, it means that $\delta = (\varphi \rightsquigarrow \neg\psi)$ is available. For the order $\leq$ we omit reflexive loops and links obtainable by transitive closure to ease readability.

Since $\delta_1 < \delta_2$ holds true, a rational agent should only endorse the default $\delta_2 = (p \rightsquigarrow \neg f)$ in the Tweety-case (Horty, 2012, pp. 23–25, 32–33); and consequently end up with $\Gamma = Th(\{p, p \rightarrow b, \neg f\})$.

To conclude our formal framework we refine our definition of a single agent default theory, enabling it to handle cases with multiple agents. A multi-agent default theory is tuple:

$$\Delta_i = (W_i, D_i, \leq_i, \Gamma_i)_{i \in A}$$

where 'A' denotes a countable set of agents with typical elements a, b, c...

2.4 The Model

Now, let's put our formal framework to use and construct a model of the **Young Earth Creationist** as advertised earlier. The agents disagreeing—i.e., Henry and Richard—are endorsing different fundamental epistemic principles (modeled as *fundamental* default rules) with incompatible conclusions. More explicitly:

multiple extensions, we can interpret every extension of a default theory as a possible equilibrium state that an ideal reasoner might arrive at—i.e., as a *possible belief state*.

- ▷ Let 'p' represent the target-proposition of the disagreement, viz., $\langle$*Planet Earth is no more than 6000 years old*$\rangle$;
- ▷ let 'q' denote the proposition $\langle$*The Bible asserts that Planet Earth is no more than 6000 years old*$\rangle$;
- ▷ and finally, let 'r' refer to the proposition $\langle$*The scientific consensus is that Planet Earth is more than 6000 years old*$\rangle$.

So, Henry endorses fundamental default $\delta = (q \rightsquigarrow p)$ whereas Richard endorses fundamental default $\delta' = (r \rightsquigarrow \neg p)$, which suggests the following three-step logical analysis of their disagreement.[13]

1. *Initial situation*. Henry and Richard's situation can be explicated using a multi-agent default theory $\Delta_i = (W_i, D_i, \leq_i, \Gamma_i)_{i \in A}$. Let '$a$' refer to Henry and '$b$' to Richard (such that $a, b \in A$). We can assume that a and b each has internally consistent belief sets, and that q is in a's background information while r is in b's ditto. Given a's endorsement of δ and b's endorsement of δ', the belief set $\Gamma_a \cup \Gamma_b$ is inconsistent (by the definition of belief set). Hence, a and b are in a state of potential deep disagreement.
2. *Appreciation*. a and b realize that they are in deep disagreement.
3. *Update*. a and b exchange information about their respective positions, thus we need an updated multi-agent default theory to capture a state of full disclosure: $\Delta'_i = (W'_i, D'_i, \leq'_i, \Gamma'_i)_{i \in A}$, where $W'_i = W_i \cup W_j, i \neq j$; $D'_i = D_i \cup D_j, i \neq j$. As D_a and D_b are disjoint (yet comparable) the ordering $\leq'_i$ can either be specified such that for all fundamental $\delta_i \in D_i$ and all fundamental $\delta_j \in D_j$: $\delta_i >'_i \delta_j$, or $\delta_i ='_i \delta_j$, or $\delta_i <'_i \delta_j$. Each of these corresponds to a specific type of response to deep disagreement.

- ▷ **Steadfastness**: $\delta_i >'_i \delta_j$ represents a conservative rationale where the new information is considered of less priority than the old. Hence, the beliefs of both agents will be unaffected by full disclosure.
- ▷ **The Equal Weight View**: $\delta_i ='_i \delta_j$ represents a rationale where new and old information is considered equal. This is a strong conciliationist rationale leading each agent to suspend judgement—i.e., each agent would become undecided about what fundamental principle to endorse after the update (on the pain of inconsistency).[14]

[13] It would actually be a fair objection to claim that realistic instances of *fundamental* epistemic principles should be captured by some much more *schematic* default rules than those suggested here. Yet we allow ourselves to neglect this complication here in order to keep things simple.

[14] This outcome is technically unproblematic for us because default theories allow for multiple extensions (cf. footnote 12).

- ▷ **World View Switching**: $\delta_i <'_i \delta_j$ represents a rationale where the new information is considered of higher priority than the old. Thus, a would adopt b's initial belief set (cf. step 1) and *vice versa*. While this response may seem unrealistic, it neatly captures the drastic nature of deep disagreement, i.e., succumbing to one's opponent means giving up one's fundamental epistemic principle(s).[15]

2.5 Discussion

So far, so good. The model we have just constructed is both provisional and extremely simple, yet it does quite well in modeling the interpersonal dynamics of typical disagreement cases. It is a first-mover in modeling *deep* disagreement—where disagreeing isn't simply a matter of having incompatible beliefs regarding surface-level propositions, but also a tension between fundamental epistemic principles—using the tools of default logic, and it can easily be augmented to bring about more sophisticated models, e.g., by drawing on technical results in Default Logic from artificial intelligence, or from neighboring fields such as AGM Belief Revision (Alchourrón et al., 1985; Hansson, 2017) and Epistemic Logic (Hintikka, 2005; Rendsvig and Symons, 2019). The model captures standard responses to disagreement known from the epistemological literature, i.e., Steadfastness (Kelly, 2005; Titelbaum, 2015) and (strong) Conciliationism (Elga, 2007; Matheson, 2009), and shows surplus by mirroring the drastic nature of deep disagreement qua World View Switching.

These merits notwithstanding it's fair to ask whether our framework of default logic does a better job modeling deep disagreement than its most obvious rivals. Consider first *Classical Propositional Logic.* In this framework one could represent fundamental epistemic principles as material conditionals. Default logic, as defined above, is merely an extension of classical propositional logic, so in case the classical framework does equally well modeling deep disagreement, Ockham's Razor would force us to adopt the classical alternative.

This move would come with a serious drawback for our present purposes, however. Notice that in our model from §2.4 we represent fundamental epistemic principles as default rules—i.e., as metalinguistic items beyond the scope of explicit evaluation. In contrast, treating such principles as material conditionals would make them part of an object-language, and thus eligible to explicit evaluation (as objects of belief). This is clearly an undesirable feature of the classical framework when it comes to modeling deep disagreement. Fundamental epistemic principles are supposed to be the kind of principles we normally take for granted in disagreements; not just ordinary targets of evaluation (Wittgenstein, 1969b; Fogelin, 2005;

[15] A potential fourth response to deep disagreement—where one side is steadfast while the other switches their world view—would require us to accept that two rational agents can react differently upon realizing deep disagreement. Whether this is a tenable option depends on our understanding of rationality, see, e.g., (Fogal and Worsnip, 2021), for a useful discussion of *structural* versus *substantive* rationality.

Lynch, 2010, 2016; Kappel, 2012). Hence, our modeling of fundamental epistemic principles as metalinguistic default rules seems superior to at least one rival.

But how about *Subjective Bayesianism*? In a Bayesian framework one could represent fundamental epistemic principles as probabilistic update functions.[16] This would enable us to model a common sense response to (deep) disagreement, which is neglected by our default logic-model, viz., adjusting one's confidence levels appropriately in the target-proposition under dispute (and in one's background assumptions and epistemic standards).

Even so, for the purposes of modeling *deep* disagreement in particular there seems to be a serious downside to Bayesianism. On the assumption that there is exactly one rational update function, the Bayesian will be unable to model *rational irresolvability*. For a disagreement to count as rationally irresolvable—by Bayesian lights—the parties involved would need to endorse non-equivalent update functions. Otherwise even agents with radically different priors would eventually converge on a rational credence. This seems to count in favor of our default logic-model because its agent-relative ranking of defaults allows two completely rational agents to disagree with each other.[17]

Summa: we have constructed an elementary model of deep disagreement using the technical tools of default logic, and compared the result with some obvious competitors. We have seen that our simple model fares quite well in comparison to both Classical Propositional Logic and Subjective Bayesianism. Of course we haven't made a decisive argument for default logic *vis-à-vis* modeling deep disagreement, but as stated, our proposed model is merely meant as a provisional one to be further discussed and refined, as is indeed the very core of the model-building perspective.

2.6 Conclusion of §2

An interesting conclusion that is revealed from our formal modeling exercise above is that *if* we accept that logical disagreements are indeed structurally similar to deep disagreements, which seems well-motivated, then how we ought to respond to them, epistemically speaking, will depend on whether it is legitimate to have an entirely subjective ranking of logical principles. In §2.5 we took the subjective ranking of defaults to be a positive feature of the Default Logic-model in comparison to Bayesianism because it allows two completely rational agents to disagree with each other. However, this very feature might turn out to be a downside of the model when it comes to logical disagreements (understood as deep disagreements). Since it seems rather undesirable to have agents rank *logical* principles in a completely subjective fashion. Philosophers usually consider it a clear desideratum of their systematic thinking

[16] See (Talbott, 2016) for more on (Subjective) Bayesianism.

[17] This is at least true on accounts that understand rationality as internal coherence.

that logical principles (and their ranking) should be objective in some sense. Take—for example—Jack Woods, who recently wrote:

> Our second intuitive feature of logic is its objectivity. Logic is not supposed to be relativized to individuals or cultures; arguments don't become correct by being articulated in different cultures; and so on. Logic is about what follows from what and why, period. (Woods, 2023, p. 32)

As implied by this passage, something has gone wrong if logical principles are modeled as agent-relative. So, while a subjective ranking of defaults might be a *pro* feature of our Default Logic-model for some tasks, e.g., modeling paradigm examples of deep disagreement like the **Young Earth Creationist**, it seems to be a *con* feature when it comes to capturing the objectivity which is usually associated with logical disagreements.

Chapter 8

Epilogue

1 The Epistemic Significance of Convergence in Logical Theorizing

We'll end with some reflections on the interesting—and yet underexplored—epistemological flip side of logical disagreement, viz., the epistemic significance of agreement (or convergence) in logic.[1] To see why this topic is intriguing, consider a case where two incompatible logical theories have a non-empty intersection including the entailment-sentence expressing the proposition $\langle$*Modus Ponens is valid*$\rangle$. Intuitively this should increase our confidence in the overlap. We should feel more confident that the proposition $\langle$*Modus Ponens is valid*$\rangle$ is true, now that this is common ground between the combatants.

Or should we? While the idea of convergence is familiar from the philosophy of science (Oberauer and Lewandowsky, 2019; Schupbach, 2022), there is almost no sustained discussion of the nature and epistemic significance of convergence in logical theorizing.

Perhaps convergence in logic is epistemically significant because converging logical theories constitute independent methods of reasoning confirming the same results, e.g., that Modus Ponens is a deductively valid inference, and when independent methods confirm the same results, we have more reason to trust them. While initially appealing, there are serious challenges to this idea upon reflection.

The first challenge lies in understanding what it means for methods of reasoning to be independent of each other on a generic level and explaining why convergence of such independent methods is epistemically significant. A second challenge is deciding how (or if at all) logical

[1] Parts of this epilogue are based on joint work with Klemens Kappel, Andreas Christiansen, and Victor Lange.

theories may be considered independent methods of reasoning, the convergence of which is epistemically significant. As will become clear, even if we can say something useful about the first of these challenges, the second is not a straightforward matter.

1.1 A Sketch of a Generic Argument

Let's start by considering Goldman's seminal work on expert trust and consensus (Goldman, 2001). Suppose two experts Albert and Bohr agree with respect to some proposition $\langle p \rangle$ within the remit of their expertise, say, physics. Normally we would consider this a reason to trust what they agree on more. But—as Goldman points out—if Bohr merely registers what Albert thinks about $\langle p \rangle$ and then blindly follows this, their agreement has no, or only very little, epistemic significance. So, clearly the epistemic significance of expert agreement depends on the experts being independent in some sense. Following Goldman, let's say that two experts S and S^* are *Probabilistically Independent* if and only if the probability that S asserts that $\langle p \rangle$, given that $\langle p \rangle$ is the case and that S^* asserts that $\langle p \rangle$, is equal to the probability that S asserts that $\langle p \rangle$, given that $\langle p \rangle$ is the case and that S^* *doesn't* assert that $\langle p \rangle$, and *vice versa* (Goldman, 2001, pp. 99-104). Clearly, if S and S^* are highly competent in their domain of expertise and probabilistically independent, their agreement is epistemically significant.

Now, it's important to note that probabilistic independence is a quite narrow notion. Suppose that expert agents Alma and Bashir use the same method M to investigate a particular question, but each does so in the privacy of their own labs, not knowing what the other does. This makes Alma and Bashir probabilistically independent, according to the definition just stated. Yet, the epistemic significance of this investigation is equal to the significance of an alternative investigation in which only one of Alma and Bashir uses M on two distinct occasions and compares the results (assuming that they are equally competent in using the relevant method). What this essentially tells us is that the probability that there has been an error in the application of M is smaller than it would have been, had M only been used once, and it reduces the influence of random variations in the results produced by M. It doesn't speak to the idea that two agents agree on some result based on different methods.

So, let's consider situations where experts are not only probabilistically independent, but also base their results on different methods, viz., *Methodological Independence*. In natural science, two methods are often seen as independent insofar as they exploit different causal mechanisms. For example, one familiar method for measuring temperature is based on the expansion of a liquid with increasing temperature, whereas another method is based on the correlation of infrared radiation and temperature. These two methods are independent because they exploit different causal mechanisms (Woodward, 2006).

Why, then, is methodological independence epistemically significant? The straightforward answer is that when methods exploit different causal mechanisms, they tend to err under different conditions. One method may fail to work reliably in certain circumstances, but some other method exploiting a different causal pathway may still work well under those conditions (think again of the temperature measuring-example given above). This fact should make us trust converging results more when they arise from independent methods: when two independent methods produce the same result, it's less likely that they have both malfunctioned, because independent methods malfunction under different circumstances.

Let's expand and generalize this idea. First, a generic formulation for methodological independence suggests itself: M_1 and M_2 are independent methods if and only if there are relevant circumstances C, where M_1 is prevented from working with its characteristic reliability, but M_2 still works with its characteristic reliability in C (and *vice versa*).

This is of course a rough characterization in many ways. Methodological independence will be a matter of degrees on several dimensions. Circumstances in which two methods differ in reliability may cover a smaller or larger domain, they may be more or less modally remote, and the differences in reliability across circumstances are also a matter of degree. However, there is an intuitively attractive explanation of the epistemic significance of independent method convergence. When we consider whether we should trust a result provided by a given method, we need to factor in the possibility that the method malfunctioned. If two independent methods converge when applied in the same circumstances, it's less likely that they both malfunction. Learning that two independent methods converge in a given instance gives us new evidence that permits us to increase our confidence that none of them have malfunctioned.

To spell this out a bit, consider here a sketch of a *Method Convergence Argument*. Assume that (1) two methods M_1 and M_2 are both highly reliable when applied in inside-domains. Assume that (2) the inside-domains of M_1 and M_2 overlap considerably, i.e., in many possible cases M_1 and M_2 are both highly reliable in the same circumstances. Yet, (3) when applied in outside-domains, M_1 and M_2 are each much more likely to malfunction. Finally, assume that (4) M_1 and M_2 are distinct in the sense that there are relevant circumstances where M_1 is prevented from working with its characteristic reliability, but M_2 still works with its characteristic reliability (and *vice versa*). So, the set of cases where M_1 is unreliable is not identical to the set of cases where M_2 is unreliable. One may suggest that when all of conditions (1)-(4) hold, we should be more confident in the result when the two methods converge. Because convergence on the correct result is more likely when both methods are working within their inside-domains than when one or both are working in outside-domains.

Note that this generic argument uses the fact of convergence as a premise in a way that encapsulates the distinct significance of convergence. So, the very fact that two methods converge figures as a reason to increase confidence in their result beyond what is given by any initial trust we might have in each of the methods individually. This identifies a sense in which convergence is epistemically significant *eo ipso*. Note also that the Method Convergence Argument doesn't rely on specific information about domains where a method might malfunction. Normally, of course, if we know how a given method works, we'll have some ideas about the conditions under which it fails. But the argument doesn't depend on such information. Even if we don't know exactly when our methods will malfunction, it is still the case that if we know that conditions (1)-(4) hold for them, and that they agree with respect to $\langle p \rangle$, then we have some reason to increase confidence in the specific $\langle p \rangle$ they agree about. Of course, the Method Convergence Argument could be refined further, but we need not do so here. The suggestion is merely that the argument above gives a rough generic account of the epistemic significance of convergence between independent methods.[2]

To really appreciate the account, which was just presented, consider an intuitive case for the epistemic significance of convergence:

> **Clockwork**. James owns two vintage clocks. One afternoon he needs to know exactly what time of the day it is, as he is expected to make an important phone call. He then turns to the two vintage clocks. Last time he had a look at them, about two weeks ago, he adjusted the clocks to follow time correctly. However, being vintage and having been stowed away in a box for the past two weeks, he knows that they may not be entirely accurate. In fact, prior to consulting the clocks, he is (for some reason) warranted in credence 0.8 that one of the clocks is fully accurate today. James looks at the first clock, it reads exactly 2pm. James consults the second clock, and it turns out that it also reads exactly 2pm. Should James now have a credence of 0.8 that the time is 2pm, as he would be entitled to by disjunctive reasoning? It seems not. Rather, his confidence that both clocks read correctly should increase considerably above 0.8, and he should have a correspondingly high credence that the time is indeed 2pm. The distinct fact that his two vintage clocks converge is significant evidence that none of the clocks have malfunctioned in the way he suspected they might have.[3]

Now, there is in fact a simple approach to calculating how much additional reliability we get from distinct methods converging. Suppose we have two distinct methods M_1 and M_2 that

[2] It's worth remarking that there are cases where the convergence of two methods warrants significant confidence, even if neither method is reliable. Suppose, for example, that two first-year history students (who are independent in the relevant way) tell you that the Dano-Swedish Scanian War was fought from 1675 to 1679. Even if we assume that both students are fairly unreliable, say, there is a 10% chance they get questions like this right, their convergence warrants high confidence in their answer, since it is much less likely that they provide any specific wrong answer, say, 1%. Since there is *ex ante* a 1% chance of convergence on the right answer, but only a 0.01% chance of convergence on a wrong answer, we have a warranted credence *ex post* of 0.99 in the convergent answer.

[3] In this case James' confidence should in fact be very close to 1 because the probability that a broken clock shows some particular time is very low (1-in-720 for each value in whole minutes, e.g., 2:01, assuming that the probability is equally distributed across each whole-minute value). The probability that both clocks are wrong in the same way is miniscule, and much smaller than the probability that at least one clock is correct.

we can apply to a fixed pool of cases. Assume that M_1 and M_2 each has some probability of returning a correct verdict when applied to a case in the pool. We can then compare various strategies. A first strategy would be to run one of the methods—M_1 for example—on all cases, and then count how many of the verdicts issued are correct. The ratio of correct verdicts will give some indication of how much we should trust the used method. A second strategy would be to apply one of the methods, say M_1, twice to each case in the pool and then count all the instances where M_1 agrees with itself with respect to a given case. We can then calculate the number of *correct* single method convergence verdicts relative to the total number of single method convergence verdicts. The higher this ratio, the more trust should we place in such "intra-method convergence." Finally, we could apply both our methods M_1 and M_2 once to each case in the pool, then count the instances where they converge on the correct verdict, and compare to the total number of convergent verdicts. What the method convergence argument suggests is that the dual-method convergence ratio of correct verdicts should be higher than the intra-method convergence ratio of correct verdicts; which in turn should be greater than the ratio of correct verdicts resulting from the use of a single method applied to each case in the pool only once.

To elaborate, suppose our pool of cases is divided as follows.

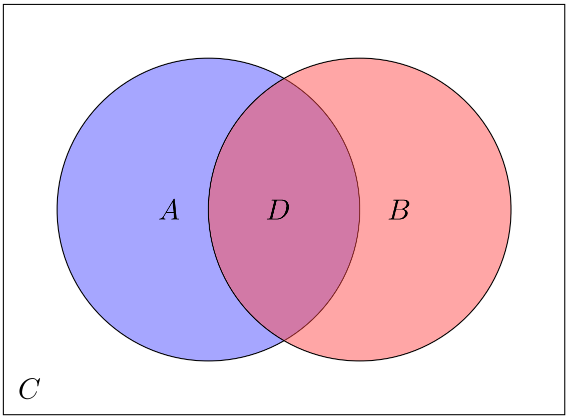


Figure 8.1: Venn Diagram

Within the pool there is a particular subset of cases where method M_1 has a high inside reliability. Outside of this subset of cases, by contrast, method M_1 has a low outside reliability. Similarly, for M_2 with respect to a different subset of cases. Now, the sets of cases where M_1 and M_2 work well, respectively, are partly overlapping. Suppose A is the set of cases that fall within the reliable domain of M_1, but outside the reliable domain of M_2. Suppose that B is the set of cases where M_2 is particularly reliable, but which is outside the reliable domain of

M_1. Let C be the set of cases outside the reliable jurisdiction of both methods, and finally say D is the set of cases where M_1 and M_2 overlap in their reliable domains.

We can now state what we are interested in more clearly. Assume that in all cases there is a truth of the matter, call it 'p'. If we apply one of the methods, say M_1, to all cases in our pool (that's all cases in A, B, C, and D), how many p-verdicts should we expect relative to the total number of verdicts? Obviously, the answer depends on the inside and outside reliability of M_1, as well as the distribution of cases in A, B, C, and D. But say we keep all this fixed. Suppose that we apply M_1 twice to each case in the entire pool, and then look only at the instances where M_1 agrees with itself. How many times will M_1 agree with itself on the verdict p, compared to the total number of agreement verdicts? Finally, suppose we apply both M_1 and M_2 once to each case in the pool. How many p-agreement verdicts should we expect, compared to the total number of agreement verdicts?

To illustrate, let's be concrete and assume that M_1 and M_2 both have an inside reliability of 0.9, whereas the outside reliability is only 0.1. Assume that A, B, and C, each contains a fraction of 0.1 of all cases. This leaves a fraction of 0.7 of all cases in the overlapping D. Using these values in a few simple calculations, we get that the truth-ratio resulting from applying one of the methods to all cases in the pool only once is 0.74. Using the same method twice on all cases in the pool to calculate the number of *correct* single method convergence verdicts relative to the total number of single method convergence verdicts, we get a slightly higher truth-ratio of 0.79. Thus, intra-method convergence is better than just using a single method once. But using our two different methods one time each on all cases in the pool to calculate the instances where they converge on the correct verdict, and compare to the total number of convergent verdicts, gives an even higher truth-ratio of 0.85 (an improvement by a factor of 1.07). So, dual-method convergence is better than intra-method convergence. When the reliability of our methods and the distribution of cases are as specified above, and kept fixed, using one method twice is better than using it only once, but using two different methods is better still. This confirms our conjecture: there's some epistemic significance to convergence. When two distinct methods agree, we have more reason to trust the result.

Bearing this concrete example in mind, we can get a better grasp of the basic features that matter for intra-method and dual-method convergence. As we saw, the epistemic significance of dual-method convergence depends on the truth-ratio, i.e., the probability that two methods converge on the correct result p over the probability that they converge *simpliciter*:

$$\frac{Pr(Convergence\,(p))}{[Pr(Convergence\,(p)) + Pr(Convergence\,(\neg p))]}$$

The decisive feature is really the following ratio:

$$\frac{Pr(Convergence\,(p))}{Pr(Convergence\,(\neg p))}$$

Thus, the key issue is the ratio between: the probability that M_1 and M_2 converge on p across A, B, C, D; and the probability that M_1 and M_2 converge on not-p in A, B, C, D. The higher this number, the more epistemically significant the convergence. We can also observe some features that potentially increases the number. The higher the reliability of M_1 and M_2, the more likely they are to converge on p. M_1 and M_2 are more likely to converge on p in D than in anywhere else. Hence, the larger the overlap in D, relative to the number of cases in A, B, and C, the more epistemic weight will convergence on p acquire. Furthermore, the ratio is sensitive to the way that M_1 and M_2 work in their respective outside domains. In the calculations above, we have assumed that methods are returning either p or not-p. This implies that a low reliability in the outside-domains is equivalent to a high probability of a not-p-verdict in such areas. For example, if a method has a low reliability of 0.1 in an outside-domain, this means that it has a 0.9 probability of returning a verdict of not-p there. While surely useful for simple calculations, the assumption that the outputs of our methods are binary in this way will only be realistic in some settings.

1.2 Is Method Convergence Epistemically Significant in Logic?

The next step is to consider if this rationale for the epistemic significance of method convergence can be applied to the convergence between logical theories. The first caveat we need to face is this: *Can we regard logical theories as (independent) methods of reasoning?*

As we have already seen multiple times in this monograph, it isn't obvious that we can. As noted, Gilbert Harman has forcefully challenged the view that logic is normative for reasoning (Harman, 1984, 1986). Deductive logic and reasoning are two fundamentally different enterprises—logical principles are not in any direct sense rules of belief revision, he argues (Harman, 1984, p. 107).

A second caveat is that logicians might converge in their views for many types of reasons, some of which would undermine rather than support the epistemic significance of convergence. For instance, we might be affected by socio-culturally embedded biases that drive us to believe that our basic theories support familiar or popular views, irrespective of what the theories actually assert. Similarly, individuals in intellectual environments might have a desire to align with others, and this—rather than properties of theories they accept—may explain convergence. These seem to be occurrences of convergence where the underlying causes are not of the right kind to insure epistemic significance.

Finally, and possibly even more concerning, is the caveat that the epistemic significance of convergence between logical theories seems to depend heavily on the ontology of logic. What is logic about? As in meta-ethics, not everyone is ready to grant that there are hard facts we can seek out in logic. *Non-cognitivists about logic* hold that there are no logical facts at all, and thus there is no fact of the matter as to whether certain entailment-claims or logical laws are true (Field, 2015). Accepting non-cognitivism or similar ontological views would have severe epistemological repercussions for the prospects of applying the generic argument to logic, since it is at best unclear how method convergence in logic could be epistemically significant against this backdrop. Perhaps two logicians reasoning from different logical traditions are not at all like two different witnesses independently observing the same crime scene from different locations and reporting the same facts, where their perceptual apparatus are independent reliable indicators of what happened. Perhaps convergence between two incompatible logical theories—still agreeing that Modus Ponens is valid—isn't like the case of the two vintage clocks agreeing on the time of day (cf. §1.1). When one learns that Modus Ponens is valid according to a theory of classical logic and accepts this, and then later learns that it is also valid by intuitionistic lights, one does in a sense get a second opinion about the validity of Modus Ponens, but arguably not one that bolsters the first in an epistemic way.

In his momentous three-volume book *On What Matters*, Parfit (2011) argues that "*...different moral theorists are climbing the same mountain from different sides.*" Alas, whether something similar is the case for logical theorists is not a less vexed issue.

Chapter 9

Appendix I

In this first of two technical appendices we'll clarify and discuss Michael G. Titelbaum's *Fixed Point Thesis* and his No Way Out-argument in support of it. We'll also see how Titelbaum relates his discussion of the Fixed Point Thesis to the Right Reasons View about peer disagreement. Appendix I is included in the monograph because Titelbaum's thesis is notoriously controversial among epistemologists, it is often misunderstood, and it would have very severe consequences for the epistemology of disagreement if it were true.

Keywords Titelbaum's Fixed Point Thesis; Rationality's Fixed Point; Peer Disagreement; the Right Reasons View; Epistemic Rationality; Rational Reflection

1 Rationality's Fixpoint

This section outlines Titelbaum's *Fixed Point Thesis* (henceforth 'FPT') and his initial motivation for it (as described in (2015)).[1] We'll also discuss some intuitive reasons against the thesis.

From the outset FPT has been wedded with controversies. In slogan form FPT claims that: Mistakes *about* the requirements of rationality are mistakes *of* rationality (Titelbaum, 2015, p. 253). This implies that any false belief about the requirements of rationality involves a mistake not only in the sense of believing something *false* but also in being distinctively *irrational* in one's reasoning. Thus it's impossible to have rational but false beliefs about what rationality requires of you in any given situation.

[1] The thesis also goes under the name 'Rationality's Fixed Point' (Titelbaum, 2015, 2019).

At first sight FPT might strike the reader as downright implausible, and Titelbaum is actually happy to admit this much. According to him, not only is FPT *prima facie* counterintuitive, it also comes with a host of surprising epistemic consequences, including that certain kinds of justification will be deemed indefeasible (contrary to the contemporary default in mainstream epistemology):

> Every agent possesses a priori, propositional justification for true beliefs about the requirements of rationality in her current situation and this justification is 'ultimately empirically indefeasible'. (Titelbaum, 2015, p. 276).

Another consequence of FPT is that misleading all-things-considered evidence about rationality requirements is impossible. Consequences such as this one have led some authors, e.g., Claire Field (2019), to deny FPT on the grounds that it is too demanding for human agents. Plausibly, we could have false but rational beliefs about any subject matter including rationality requirements. Why should beliefs that are specifically about what rationality requires of us have special status? To fully grasp Field's critique of FPT, let's consider an example.

> **Science**. Suppose that an otherwise incredibly reliable source tells me that I ought to accord my beliefs with the truths of science X in order to be perfectly rational. Following this advice, I believe that rationality requires me to do so, but this happens to be false (by assumption).

In this (and similar) situations it seems epistemically rational for me to hold a false belief about what rationality requires of me. After all, my belief is based on a credible piece of testimony even if the information I received is false; and note that even rationally ideal agents could end up in similar circumstances. If, say, an epistemically ideal agent is told by an incredibly reliable informant that they have ingested a pill making their reasoning skills non-ideal and prone to errors that are undetectable from the inside, and yet this piece of testimony happens to be false. It seems that this agent should have a "less than certainty"-confidence that rationality requires a higher-order credence of 1 regarding their own epistemic idealness (even if certainty about these matters is in fact what rationality requires of the agent in these circumstances).[2]

It should also be noted that an important qualification—which is entirely absent from FPT in its slogan form—is that Titelbaum restricts the jurisdiction of FPT to *a priori* rationality requirements (Titelbaum, 2015, p. 254). With this in mind, we can consider FPT in regimented form:

[2] For similar cases concerning highly idealized agents, see (Bradley, 2021).

FPT No [epistemic] situation rationally permits an a priori false belief about which overall [doxastic] states are rationally permitted in which situations. (Titelbaum, 2015, p. 261).

For an illustration of how this version of FPT is supposed to work we can ponder the well-known logical omniscience requirements from formal epistemology. Standard Bayesian epistemology takes Kolmogorov's probability axioms to represent rational requirements on agents' degrees of belief (Talbott, 2016). One such axiom (also known as 'the Normality Axiom') tells us to assign value 1 to every logical truth (Schupbach, 2022). Now, suppose that an agent S believes that standard Bayesianism provides genuine rationality requirements and that rationality requires her to have a credence of 1 in every logical truth, but that S in fact fails to do so with respect to a given logical truth φ (perhaps an extremely complicated logical truth). Because S's $Credence(\varphi) < 1$ in this case, it follows from FPT that S's overall doxastic state is irrational—and this will look like a completely outrageous verdict to some epistemologists.

1.1 Rational Reflection

At this point it would only be understandable if some readers are quite curious about Titelbaum's initial motivation for FPT. What in the world could have driven Titelbaum to defend and endorse this radical thesis to begin with? As it turns out, Titelbaum motivates FPT by appealing to a specific kind of *rational reflection*:

> Every agent possesses a priori, propositional justification for true beliefs about the requirements of rationality in her current situation. An agent can *reflect* on her situation and come to recognize facts about what that situation rationally requires. Not only can this *reflection* justify her in believing those facts; the resulting justification is also empirically indefeasible (Titelbaum, 2015, p. 276) (our emphasis).

Now, while this reflection might have a highly idealized ring to it—and though Titelbaum admits that he first came to think about FPT while studying the problem of logical omniscience (Titelbaum, 2015, p. 257)—he nonetheless proclaims that the thesis concerns all-things-considered evaluations of "attitudes held by actual agents," not just perfectly rational ones (Titelbaum, 2015, p. 288).[3] Thus, Titelbaum isn't ready to grant critics like Field (2019) that FPT is too demanding for human agents.

In order to firmly understand where Titelbaum's appeal to reflection is coming from, the reader should notice how Gödel's famous work involving fixpoints of logical (and mathe-

[3] For a recent contribution to the debate on *ideal* versus *non-ideal* epistemology, see (Carr, 2021). See also (McKenna, 2023).

matical) theories can provide a useful formal analogy to Titelbaum's informal version of rational reflection. In particular, one can appeal to Gödel's Fixpoint Lemma also known as a "*reflection principle*" stating that $T \vdash Prov_T(\ulcorner\varphi\urcorner) \leftrightarrow \varphi$ for an illustration (cf. (Zach, 2019)). According to Gödel's reflection principle a formal theory T can "reflect" on what it can derive and come to recognize not only what is true in T but also what is *provable* in T. In this sense, T can reflectively move back and forth between provability and truth. Similarly, on Titelbaum's view, suppose for instance that a given subject S is rationally committed to the axioms of Peano Arithmetic (cf. Chapters 5 & 6), PA, then if S reflects on these axioms, they can come to recognize what rationality requires given their PA-commitments, and come to believe (with empirically indefeasible justification according to Titelbaum) not only that $\langle 2+2=4\rangle$, but also that $\langle$it's rationally required for S to believe that $2+2=4\rangle$. This does indeed sound a lot like what you get with a provability predicate, moving from $\langle 2+2=4\rangle$ to $\langle$it's provable in PA that $2+2=4\rangle$.

Notice finally how strikingly insignificant the role of the thinking subject S is in Titelbaum's appeal to reflection. FPT is a thesis concerning *propositional justification* rather than *doxastic justification*, i.e., the justification of propositions rather than belief-tokens (cf. Chapters 1 & 2). Doxastic justification is a property that a belief has when one believes a proposition for which one has propositional justification, and this belief is based on that which propositionally justifies it. So, in the context of FPT, it doesn't matter for the justificational status of S's overall doxastic state whether S actually goes through some reflective process and ends up realizing what rationality in fact requires in the given situation. What matters according to FPT is that S could—in principle—do so.

2 Titelbaum's No Way Out-Argument

Let's next turn to the so-called "No Way Out"-Argument for FPT, which we take to be Titelbaum's primary argument for his thesis. The argument takes a general akrasia-constraint on epistemic rationality as its premise and proceeds deductively.[4] Otherwise put, Titelbaum expects FPT to follow from a premise already accepted by most, viz., that states of akrasia are irrational (Titelbaum, 2015, pp. 253-254). Recall once more the Akratic Principle ('AP') from Chapter 6:

AP No [epistemic] situation rationally permits any overall [doxastic] state containing both an attitude A and the belief that A is rationally forbidden in one's situation. (Titelbaum, 2019, p. 227)

As we know by now AP is usually taken to imply that you should either have the attitudes you

[4] A different argument for FPT can be found in (Littlejohn, 2018).

believe you ought to have, or stop believing that you ought to have those attitudes. Hence, in the name of rationality, AP imposes two distinct rationality constraints on the combinations of attitudes one can have while being epistemically rational:

> Example of epistemic akrasia (mistake of type 1): Anna believes that $\langle p \rangle$ while also believing that $\langle$*believing p is rationally forbidden in her current situation*$\rangle$;
>
> Example of epistemic akrasia (mistake of type 2): Anna fails to believe that $\langle$*it is raining*$\rangle$ while believing that $\langle$*believing that* it is raining *is rationally required in her current situation*$\rangle$.

To grasp Titelbaum's specific use of the term 'rational' and his argument for FPT, we introduce the following notation (inspired by works on a *rule-based* approach to rationality due to John Broome (2007) and Mattias Skipper (2019)).[5] Define R, a (total) *Rational State Function*

$$R : \mathcal{S} \mapsto \mathscr{P}(D)$$

taking us from possible epistemic situations in $\mathcal{S}$ to sets of doxastic states in D (where a doxastic state corresponds to a set of individual doxastic attitudes toward different propositions).[6] Let '$R(s)$' denote the set of doxastic states that someone who is in a particular situation s is rationally permitted to be in. A simplifying assumption about R is that we assume a tripartite view of doxastic attitudes rather than a more fine-grained view (exactly as Titelbaum does, cf. footnote 6 below).

In line with the above definitions we can then define:

> *Rational Permission*: Doxastic attitude A is rationally permitted in epistemic situation s if and only if $A \in d$, for some doxastic state $d \in R(s)$.
>
> *Rational Requirement*: Doxastic attitude A is rationally required in epistemic situation s if and only if $A \in d$, for all doxastic states $d \in R(s)$.

We say that if A is not rationally permitted in epistemic situation s, then A is *rationally forbidden* in s. A final simplifying assumption about R is that whenever a doxastic attitude is

[5] For Titelbaum's own presentation, see (Titelbaum, 2015, pp. 227-228).

[6] Note that while Titelbaum wants to stay neutral with respect to "the true theory of rationality," he seems committed to a kind of holism which evaluates the rationality of overall states rather than individual doxastic attitudes, as his framework evaluates individual doxastic attitudes only indirectly. Further, he restricts his focus to the familiar tripartite account of doxastic attitudes, where the possible references of the term 'doxastic attitude' is exhausted by: *belief that* $\langle p \rangle$; *disbelief that* $\langle p \rangle$; and *suspension of judgement with respect to* $\langle p \rangle$ (Titelbaum, 2015, pp. 258-259).

rationally required, it is also rationally permitted (i.e., every epistemic situation rationally permits at least one (non-empty) doxastic state. As an example of a rule that governs rationality, and which uses the terminology we've just introduced, consider:

> *Perception Rule.* If an agent's situation s includes a perception that $\langle p \rangle$, then all the overall doxastic states rationally permissible to that agent in s include the belief that $\langle p \rangle$.

Of course the reader need not accept this rule as a genuine one, it is merely meant to illustrate our use of the above terminology.

2.1 From AP to SCT

With this, the last ingredient needed to spell out Titelbaum's No Way Out-Argument for FPT is his lemma the *Special Case Thesis* ('SCT'):

SCT There do not exist an attitude A and an epistemic situation s such that: (i) A is required in the situation s, i.e., $A \in d$, for all doxastic states $d \in R(s)$, and yet; (ii) it is rationally permissible in that situation s to believe that $\langle A$ is rationally forbidden in $s \rangle$ (Titelbaum, 2015, p. 268) (our notation).

Titelbaum's general argumentative strategy is to show that SCT follows from AP (by reductio), and that FPT is nothing but a logically stronger version of SCT (i.e., we can get FPT from SCT by generalizing the latter in two distinctive ways):[7]

> *Titelbaum's Reductio Argument for SCT.*[8]
>
> Suppose (for reductio) that there is a situation s and a doxastic attitude A such that A is rationally required by s, and yet s also permits an overall state containing the belief that A is rationally forbidden, i.e., suppose that SCT is false. If $A \notin R(s)$, then we immediately have that both $A \in R(s)$ and $A \notin R(s)$ by the assumption that A is required by s (cf. the definition of Rational Requirement). If $A \in R(s)$, then we have that both $A \in R(s)$ and the belief that $\langle A \notin R(s) \rangle$ by condition (ii) of SCT. So, by AP, it follows that $A \notin R(s)$. Hence, in any case, $\perp$.

[7] SCT looks superficially similar to AP, but they are in fact logically distinct (Titelbaum, 2015, p. 267).

[8] We use the symbol '$\perp$' to express *falsum*, i.e., an atomic sentence which is always false (Restall and Standefer, 2023, p. 35).

2.2 From SCT to FPT

To complete Titelbaum's No Way Out-argument for FPT we would need to generalize SCT in two ways: (1) to mistakes other than believing that something required is forbidden; and (2) to mistakes about what rationality requires in situations other than the agent's current situation (Titelbaum, 2015, pp. 269-270).

In support of (1) Titelbaum writes:

> This generalization is fairly easy to argue for, on the grounds that any well-motivated, general epistemological view that rationally permit[s] agents to have a belief at odds with the true requirements of rationality in this direction would permit agents to make mistakes in the other direction as well. (2015, p. 270)

In support of (2) Titelbaum writes:

> Generally, an agent's total evidence will never all-things-considered support an a priori falsehood about the rational rules, because the rational rules are structured such that no situation permits or requires a belief that contradicts them. (2015, p. 274)

The reader doesn't have to accept any of (1) and (2), of course, but for the rest of this appendix we'll simply assume that Titelbaum's inference from SCT to FPT is successful for the sake of argument; and we'll focus our attention on the potential consequence of FPT for the epistemology of disagreement instead.

3 Peer Disagreement, FPT, and the Right Reasons View

Titelbaum himself relates FPT to the topic of peer disagreement. In fact he asserts that the best objection to FPT that he is aware of concerns FPT's consequences for peer disagreement cases. Titelbaum writes:

> To fix a case before our minds, let's suppose Greg and Ben are epistemic peers in the sense that they're equally good at drawing rational conclusions from their evidence. Moreover, suppose that as part of their background evidence Greg and Ben both know that they're peers in this sense. Now suppose that at t_0 Greg and Ben have received and believe the same total evidence E relevant to some proposition h, but neither has considered h and so neither has adopted a doxastic attitude toward it. For simplicity's sake I'm going to conduct this

> discussion in evidentialist terms (the arguments would go equally well on other views), so Greg's and Ben's situation with respect to h is just their total relevant evidence E. Further suppose that for any agent who receives and believes total relevant evidence E, and who adopts an attitude toward h, the only rationally permissible attitude toward h is belief in it. Now suppose that at t_1 Greg realizes that E requires believing h and so believes h on that basis, while Ben mistakenly concludes that E requires believing $\sim h$ and so (starting at t_1) believes $\sim h$ on that basis[...] At t_1 Greg and Ben have adopted their own attitudes toward h but each is ignorant of the other's attitude. At t_2 Greg and Ben discover their disagreement about h. They then have identical total evidence E', which consists of E conjoined with the facts that Greg believes h on the basis of E and Ben believes $\sim h$ on the basis of E. The question is what attitude Greg should adopt toward h at t_2. (Titelbaum, 2015, pp. 282-283)

To make the case even more concrete, Titelbaum stipulates that the initial evidence E in the Greg-and-Ben example is such that E entails $\langle h \rangle$. It could, for example, be a case involving mental math as in **Restaurant**. Further, Titelbaum assumes that Ben's mistaken conclusion that the negation of $\langle h \rangle$ is entailed by E is a genuine mistake of rationality.

Now, in response to the Greg-and-Ben case Titelbaum considers the Equal Weight View (or the 'Split the Difference view' as he calls it (Titelbaum, 2015, p. 283)), which we are already acquainted with from previous chapters (cf. the Introduction & Chapter 7), and he compares it his own favored *Right Reasons View* (cf. the Introduction). According to the Equal Weight View it follows that Greg should "split the difference" and suspend judgement about h upon discovering peer disagreement with Ben at time t_2 (assuming the familiar tripartite view of doxastic attitudes). In contrast, the Right Reasons View tells us that since Greg was rationally required to believe $\langle h \rangle$ based on E (via entailment) before discovering the relevant peer disagreement with Ben, it would be a genuine mistake of rationality if he were to give up on this believe after the discovery at t_2. But to some authors from the peer disagreement debate, the Right Reasons View gets things completely wrong here; and does so in various different ways, e.g., by neglecting the force of the higher-order evidence generated by the peer disagreement itself.

3.1 The Crowdsourcing Argument For/Against the Right Reasons View

Ironically enough, Titelbaum thinks that a good argument *for* the Right Reasons View can be developed from the basis of what he takes to be the best argument *against* it. To see how this strategic move is supposed to work, let's abbreviate the Right Reasons View by the label 'RR' (as Titelbaum does) and consider the following passage:

> Suppose for *reductio* that RR is correct and Greg shouldn't change his attitude toward h in light of the information that his peer reached a different conclusion from the same evidence. Now what if Ben was an epistemic superior to Greg, someone who Greg knew was much better at accurately completing arithmetic calculations? Surely Greg's opinion about h should budge a bit once he learns that an epistemic superior has judged the evidence differently. Or how about a hundred superiors? Or a thousand? At some point when Greg realizes that his opinion is in the minority amongst a vast group of people who are very good at judging such things, rationality must require him to at least suspend judgment about h. But surely these cases are all on a continuum, so in the face of just one rival view—even a view from someone who's just an epistemic peer—Greg should change his attitude toward h somewhat, *contra* the recommendation of RR. (Titelbaum, 2015, pp. 283-284)

Titelbaum calls this reductio the 'Crowdsourcing Argument' against RR. The gist of the argument is that there exists some number of epistemic superiors whose disagreement with Greg would make it rationally obligatory for him to suspend judgement as to whether $\langle h \rangle$; and further that there is a (much) larger number of superiors whose disagreement would make it rationally required for Greg to believe the negation of $\langle h \rangle$. We can suppose that such a transition from being rationally obligated to believe $\langle h \rangle$ to being rationally required to believe its negation happens in just two temporal steps, i.e., t_1 and t_2, and with "nice" round numbers, say, 100 and 1000 epistemic superiors, respectively.

Initially, the Crowdsourcing Argument looks like bad news for RR, but according to Titelbaum the argument proves too much. He writes:

> By supposition, E entails h and therefore rationally requires belief in it. When the experts convince Greg that E entails $\sim h$, they thereby convince him that he was required to believe $\sim h$ all along—even before he encountered them. By the Fixed Point Thesis, Greg is now making a rational error in believing that E rationally requires belief in $\sim h$. So it is not rational for Greg to respect the experts in this way. By the continuum idea, it's not rational for Greg to suspend judgment in the face of fewer experts to begin with, or even to budge in the face of disagreement from Ben his peer. We now have an argument from the Fixed Point Thesis to the Right Reasons view about peer disagreement. (Titelbaum, 2015, pp. 285-286)

Before ending this short tour through Titelbaum's work on FPT and its potentially important relationship to the peer disagreement debate, we'll make two critical observations regarding the specifics of the Crowdsourcing Argument. First, Titelbaum assumes a direct and unproblematic link between logical entailment and epistemic rationality; this much is clear from the passage "[...] E entails h and therefore rationally requires belief in it [...]" above, where 'it' refers to proposition $\langle h \rangle$. Yet, as we have seen in previous chapters—e.g., in Chapter 1—this assumption is a controversial one in the philosophy of logic since it requires a plausible *bridge principle*. Without having such a principle in place, the connection between logic and epistemic norms remains unclear. Second, Titelbaum assumes that Greg interpreted the evidence correctly in his disagreement with Ben, but one might reasonably ask: *How can Greg know that he got it right?* The answer is that Titelbaum takes RR to provide a *conditional* norm stating what an agent is rationally obligated to do upon encountering peer disagreement *if* he in fact drew the conclusion required by the evidence prior to the encounter (Titelbaum, 2015, p. 287). This aspect of Titelbaum's argumentation highlights the importance of drawing the distinction between *evaluations* and *directives* (cf. Chapter 1). RR is merely an *evaluation* of peer disagreements; not a directive which offers normative guidance from a first-person perspective.

Chapter 10

Appendix II

This final appendix provides the technical definitions of the default logic from *Reasons as Defaults* by John F. Horty (2012). The appendix is included since we relied heavily on parts of Horty's formal framework in building our model of deep disagreement in Chapter 7. Importantly, the appendix makes it transparent to the reader where our refined Horty-style default logic—as defined in Chapter 7—differs from Horty's original.

Keywords Default Logic; Reasons as Defaults

1 The Default Logic of *Reasons as Defaults*

Default reasoning follows patterns like *in the absence of reasons to the contrary, from φ, conclude ψ.* Such inference patterns are widely used in everyday thinking and are established research topics in computer science. Topics of interest in *default logic* include, e.g., what beliefs an ideal reasoner should hold given an acceptable set of *default rules* (or simply *defaults*), how some defaults defeat others, and what sets of defaults may reasonably be held jointly. A default rule may be thought of as a *defeasible generalization*, where learning the premise (by default) warrants a belief in the conclusion.

The first system of default logic was proposed and developed by Raymond Reiter (1980).[1]

[1] Reiter's default logic was one of many non-monotonic reasoning frameworks developed in the late 1970 and 80s, with early papers collected and presented in (Ginsberg, 1987), including (Reiter, 1980) and other default logic approaches (Reiter and Criscuolo, 1981; Etherington and Reiter, 1983; Touretzky, 1986; Poole, 1988), but also, e.g., circumscription (McCarthy, 1980) and modal logical approaches (McDermot and Doyle, 1980; Moore, 1985). Later, AGM belief revision theory has been proposed as a non-monotonic system (Makinson and Gärdenfors, 1991). For overviews, see, e.g., (Ginsberg, 1987; Antoniou, 1999; Delgrande et al., 2004; Antonelli, 2005; Koons,

The purpose of Reiter's seminal work was to formalize reasoning with default assumptions, to which end he used defaults of the form

$$A : C/B \tag{10.1}$$

read by Horty (2007a) as "if A belongs to the agent's stock of beliefs, and C is consistent with these beliefs, then the agent should believe B as well" (p. 386). In a Reiter default like (10.1), A is called the *prerequisite*, C is the *justification* and B is the *consequent*.[2] A Reiter default in which the justification is logically equivalent to the consequent is called *normal*. Throughout this chapter, we focus on normal defaults $A : B/B$, which we write '$A \rightsquigarrow B$'.[3]

As reasoning with defaults may be non-monotonic, classical logic does not suffice[4] as a guide for what conclusions to draw given some background information and a set of defaults (jointly called a *default theory*). Thus, a main task in default logic is to specify what conclusions are reasonable—to find the so-called *extensions* of a given default theory. Such extensions are often considered as rational fixed points that may be understood as cognitive equilibria of an ideal reasoner, and so be equated with rational beliefs held on the basis of the default theory.

To define the rational belief set(s) of an agent in an informational context, Horty's framework involves a host of notions defined in the following subsections. A rough outline of the framework is as follows.

The main aim is to establish an agent's *full belief set* given a context, i.e., a *default theory*. A default theory represents the initial data an idealized agent uses as a basis for reasoning (Horty, 2012, p.22).

Horty works with prioritized default theories, each containing a set of *hard background information* W, a set of *defaults D,* and an *order* $<$ on the defaults. How the context is arrived at is not in question, only what to believe on the background of it.

As the defaults D may produce conflicts (inconsistency), as some defaults may defeat other through priority, and as untriggered defaults should be excluded from influencing beliefs, the agent must select a reasonable subset of D on which to base their beliefs: They must find a *proper scenario.* Defining proper scenarios, i.e., scenarios that are also *rationally acceptable* (e.g., it should not allow us to conclude a contradiction from true premises), is the main task of the framework.

2017; Strasser and Antonelli, 2019). Note further that parts of the present appendix is based on joint work with Rasmus K. Rendsvig.

[2] More generally, a Reiter default may have multiple justifications, i.e., be of the form $\varphi : \psi_1, \ldots, \psi_n \setminus \chi$ given the reading "if φ is derived, and $\psi_1, \ldots, \psi_n$ are separately consistent with what is derived, then infer χ" (Brewka and Eiter, 2000), or "if φ is known and consistent with assumptions $\psi_1, \ldots, \psi_n$, then conclude χ" (Antoniou, 1999). Additionally, the formulas are normally allowed to be first-order.

[3] We still treat normal defaults as (defeasible) inference rules, and not formulas, but find the notation '$\rightsquigarrow$' easy to read.

[4] Monotonic logics—such as classical logic—satisfy that for any formula φ from the language L, if $\varphi \in L$ is a consequence of a set of formulas $\Gamma \subseteq L$ and if $\Gamma \subseteq \Delta \subseteq L$, then φ is also a consequence of Δ. That is, adding premises does not remove conclusions. Non-monotonic logics lack this property, and allow conclusions to be withdrawn in the light of new information.

Finally, the rational belief set(s) are determined: A set of formulas is a rational belief set if it is the set of logical consequences of the combination of the background information and a proper scenario. As a default theory may admit multiple proper scenarios, each may also admit multiple rational belief sets, called *extensions*.

In the following, we present the formal details of the *Reasons as Defaults* framework, with a running commentary on interpretation.

The definitions below are labeled with references to (Horty, 2012). The labels are meant as conjunctions, so for example [Def. 7, p. 17, Def. 9] specifies a definition which is based on Horty's Definition 7, content from page 17 and Definition 9. We use notation that slightly differs from Horty's (e.g., the symbol '$\rightsquigarrow$' used in default rules) and introduce a few sets (e.g., $\mathcal{D}$ as the set of all default rules), but make no alterations to concepts defined in (Horty, 2012).

1.1 Language and Default Rules

[pp. 15–18] Throughout, fix a countable set of atomic propositions Φ and a language $\mathcal{L}$ given by

$$\varphi := p \mid \top \mid \neg\psi \mid \psi \rightarrow \psi'$$

where the symbol '$\rightarrow$' denotes material implication. The remaining Boolean connectives are defined as usual. For $\Gamma \subseteq \mathcal{L}$, $\varphi \in \mathcal{L}$, write $\Gamma \vdash \varphi$ when φ is classically deducible from Γ. Denote the logical closure of Γ by $Th(\Gamma) := \{\varphi : \Gamma \vdash \varphi\}$.

Where $\varphi, \psi \in \mathcal{L}$, a **default rule**[5] is any expression of the form

$$(\varphi \rightsquigarrow \psi)$$

Denote the set of all default rules by $\mathcal{D}$ with typical elements δ, δ'. For a default rule $\delta = (\varphi \rightsquigarrow \psi)$ or a set of default rules $D \subseteq \mathcal{D}$, let

$$\begin{aligned} Premise(\delta) &:= \varphi & Premises(D) &:= \{Premise(\delta) : \delta \in D\}. \\ Conclusion(\delta) &:= \psi & Conclusions(D) &:= \{Conclusion(\delta) : \delta \in D\}. \end{aligned}$$

Intuitively, a default rule may be thought of as a *defeasible generalization*. A classic example is (*Tweety is a bird* $\rightsquigarrow$ *Tweety can fly*). By default, learning the premise warrants a belief in the conclusion, but additionally learning that Tweety is a penguin delegitimizes it. Hence, the rule is defeasible. As additional information may invalidate the conclusion, the inference

[5] Default rules are not expressible in $\mathcal{L}$, and, as in (Horty, 2012), '$\rightsquigarrow$' cannot be nested.

is an example of non-monotonic reasoning. Horty interprets defaults as providing reasons for conclusions.[6]

1.2 Default Theories

The next definition specifies the core notions of the framework: a (fixed priority) default theory represents the initial data that an idealized agent can use as a basis for reasoning (Horty, 2012, p.22). Ensuing definitions provide refinements.

[Def. 1, p.22] A **(fixed priority) default theory** is a tuple $\Delta = (W, D, <)$, where $W \subseteq \mathcal{L}$ is a set of *background information*, $D \subseteq \mathcal{D}$ is a set of *available default rules*, and $<$ is a strict partial *priority order* on D (i.e., $<$ is transitive and irreflexive).

A **scenario** based on Δ is a subset $S \subseteq D$.

Intuitively, "[...] a scenario is supposed to represent the particular subset of available defaults that have actually been selected by the reasoning agent as providing sufficient support for their conclusions—the particular subset of defaults, that is, to be used by the agent in extending the initial information from W to a full belief set, which we can then speak of as the belief set that is generated by the scenario" (Horty, 2012, p. 23).

Concerning the requirements on the relation $<$, Horty argues that transitivity is a natural requirement, that the relation should be irreflexive (i.e., strict) so that "no default can ever have a higher priority than itself" (*ibid.*, p. 20), and that the relation should not be strongly connected[7] as—though this would help to resolve conflicts between defaults—the requirement would be unreasonable, because: (1) some defaults are simply incommensurable, and (2) some defaults may have equal priorities.

Note that reason (2) contrasts with the choice of a *strict* order and suggests using a preorder $\leq$ instead. The order is then *reflexive* instead of irreflexive, with the also natural interpretation that every default is comparable to itself, and to itself it has *the same* priority. As a preorder, it may still be partial, in accordance with Horty's intuitive examples (*ibid.*, p. 20).

Horty's fixed priority default theories may be seen as a generalization of *normal* Reiter default theories, i.e., Reiter default theories (W, D) where all defaults are normal, with (W, D) represented by $\Delta = (W, D, \emptyset)$, cf. (Horty, 2007a).

[6] Horty (2007a, p. 368) writes: "Where A and B are formulas from the background language, we then let $A \rightsquigarrow B$ represent the *default rule* that allows us to conclude B, by default, whenever it has been established that A. It is most useful, I believe, to think of default rules as providing *reasons* for conclusions." In the quote, we have replaced Horty's notation '$\rightarrow$' with the present '$\rightsquigarrow$'.

[7] The priority order should not be assumed *connex,* that for any defaults δ, δ', either $\delta < \delta'$ or $\delta' < \delta$.

1.3 Proper Scenarios

Horty remarks that belief sets based on arbitrary scenarios are unsatisfactory (*ibid.*, p. 23). Satisfactory belief sets are obtained only from *proper scenarios*. The definition of a proper scenario requires the auxiliary notions of *triggered*, *conflicted*, and *defeated* defaults.

[Def. 2, p. 25, Def. 3, p. 27, Def. 4, p. 29] Let $S \subseteq D$ be a scenario based on $\Delta = (W, D, <)$. Define

$$\begin{aligned} Triggered(\Delta, S) &= \{\delta \in D \colon W \cup Conclusions(S) \vdash Premise(\delta)\}. \\ Conflicted(\Delta, S) &= \{\delta \in D \colon W \cup Conclusions(S) \vdash \neg Conclusion(\delta)\}. \\ Defeated(\Delta, S) &= \{\delta \in D \colon \exists \delta' \in Triggered(\Delta, S) \text{ such that} \\ &\qquad \delta < \delta' \text{ and } W \cup \{Conclusion(\delta')\} \vdash \neg Conclusion(\delta)\}. \end{aligned}$$

Using these three notions, Horty presents two definitions of a proper scenario. The first definition relies on the notion of a *binding default*. It is preliminary, but used throughout the book. The second is presented in his Appendix A.1 to handle certain problem cases.[8] We state the definitions in turn. [Def. 5, p. 30] Let $S \subseteq D$ be a scenario based on $\Delta = (W, D, <)$. Define

$$Binding(\Delta, S) = (Triggered(\Delta, S) - Conflicted(\Delta, S)) - Defeated(\Delta, S).$$

A scenario $S \subseteq D$ based on $\Delta = (W, D, <)$ is **stable** if and only if $S = Binding(\Delta, S)$. The scenario S is **proper**$_1$ if and only if it is stable.

The second definition is stronger, in that it implies stability, cf. Horty's Theorem 1 (*ibid.*, p. 223). It is based on the notion of an *approximating sequence*: [Def. 26, Def. 27, pp. 222–223] Let $S \subseteq D$ be a scenario based on $\Delta = (W, D, <)$. Then $(S_n)_{n \in \mathbb{N}} = S_0, S_1, S_2, \ldots$ is an **approximating sequence** based on Δ and constrained by S if and only if

$$\begin{aligned} S_0 &= \emptyset, \\ S_{i+1} &= \{\delta : \delta \in Triggered(\Delta, S_i), \delta \notin Conflicted(\Delta, S), \delta \notin Defeated(\Delta, S)\} \end{aligned}$$

The scenario S is **proper**$_2$ if and only if $S = \bigcup_{i \geqslant 0} S_i$ for some approximating sequence $(S_n)_{n \in \mathbb{N}}$.

[8] Horty exemplifies: Let $\delta = \varphi \rightsquigarrow \varphi$ and $\Delta = (W, D, <)$ with $W = \emptyset$, $D = \{\delta\}$ and $< = \emptyset$. Then $S = \{\delta\}$ is stable as δ is triggered, and neither conflicted nor defeated. Yet the belief set $E = Th(\{\varphi\})$ is not grounded in the background information.

1.4 Extensions and Beliefs

Finally, Horty defines extensions of default theories: [Def. 8, p. 32] The set E is an **extension** of $\Delta = (W, D, <)$ if there is some proper$_{\{1,2\}}$ scenario S such that

$$E = Th(W \cup Conclusion(S)).$$

This concludes the formal framework.[9]

Horty *does not* directly associate extensions with beliefs, cf. his discussion on pp. 34–40: A default theory Δ may have multiple or no extensions, and identifying the Δ-beliefs with *the* extension of Δ is therefore not well-defined. Horty discusses both multiple and lacking extensions, but he does not give a solution. As lacking extensions didn't play any role for us in Chapter 7 above, we simply ignored that problem. For multiple extensions, we conformed our terminology to what we considered the least committing of Horty's three proposals: We interpreted every extension of a default theory as a possible equilibrium state that an ideal reasoner might arrive at—as a *possible belief state*.

[9] Horty revises the definition of defeat in Chapter 8 of *Reasons as Defaults*, but writes "[...] [T]his preliminary definition is adequate for a wide variety of ordinary examples, and in order to avoid unnecessary complication, we will rely on it as our official definition throughout the bulk of this book." (Horty, 2012, p.30). The revision affects the definitions of binding defaults and of approximating sequences, resulting in the two additional definitions of proper scenarios, but neither are essential to our use of default logic. For completeness, we include the revised definition: [Def. 21, p.196] Let $S \subseteq D$ be a scenario based on $\Delta = (W, D, <)$. Define

$Defeated(\Delta, S) = \{\delta \in D\colon$ there is a set $D' \subseteq Triggered(\Delta, S)$ such that (1) $\delta < D'$, and (2) there is a set $S' \subseteq S$ such that (a) $S' < D'$, (b) $W \cup Conclusion((S - S') \cup D')$ is consistent, and (c) $W \cup Conclusion((S - S') \cup D') \vdash \neg Conclusion(\delta)\}$.

References

Adler, J. E. (2002). Akratic Believing? *Philosophical Studies*, 110(1):1–27.

Alchourrón, C. E., Gärdenfors, P., and Makinson, D. (1985). On the Logic of Theory Change: Partial meet Contraction and Revision Functions. *The Journal of Symbolic Logic*, 50(2):510–530.

Alston, W. P. (1989). *Epistemic Justification: Essays in Theory of Knowledge*. Cornell University Press.

Andersen, F. J. (2020). Uniqueness and Logical Disagreement. *Logos & Episteme*, 11(1):7–18.

Andersen, F. J. (2023a). Countering Justification Holism in the Epistemology of Logic: The Argument from Pre-Theoretic Universality. *Australasian Journal of Logic*.

Andersen, F. J. (2023b). Uniqueness and Logical Disagreement (Revisited). *Logos & Episteme*.

Andersen, F. J. (202X). Logical Akrasia. *Episteme*.

Antonelli, A. (2005). *Grounded Consequence for Defeasible Logic*. Cambridge University Press.

Antoniou, G. (1999). A Tutorial on Default Logics. *ACM Computing Surveys (CSUR)*, 31(4):337–359.

Antoniou, G., O'Neill, T., and Thurbon, J. (1996). Studying Properties of Classes of Default Logics: Preliminary Report. Springer.

Arbeiter, S. (2023). Validity as a Thick Concept. *Philosophical Studies*.

Ayer, A. J. (1952). *Language, Truth and Logic*. Dover.

Baader, F. and Hollunder, B. (1993). How to Prefer More Specific Defaults in Terminological Default Logic. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 1993)*, pages 669–674.

Baader, F. and Hollunder, B. (1995). Priorities on Defaults with Prerequisites, and their Application in Treating Specificity in Terminological Default Logic. *Journal of Automated Reasoning*, 15(1):41–68.

Bacon, A. (2013). Non-Classical Metatheory for Non-Classical Logics. *Journal of Philosophical Logic*, 42(2):335–355.

Barker, S. (2023). Deep Disagreement, Epistemic Norms, and Epistemic Self-Trust. *Episteme*.

Bealer, G. (1998). Intuition and the Autonomy of Philosophy. In DePaul, M. and Ramsey, W., editors, *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, pages 201–240. Rowman & Littlefield.

Beall, Jc. (2011). *Spandrels of Truth*. Oxford University Press.

Beall, Jc. (2017). There is No Logical Negation: True, False, Both, and Neither. *The Australasian Journal of Logic*, 14(1).

Beall, Jc. (2019). On Williamson's New Quinean Argument against Nonclassical Logic. *The Australasian Journal of Logic*, 16(7):202–230.

Beall, Jc. and Restall, G. (2000). Logical Pluralism. *Australasian Journal of Philosophy*, 78(4):475–493.

Beall, Jc. and Restall, G. (2006). *Logical Pluralism*. Oxford University Press.

Becker Arenhart, J. R. (2022a). Abductivism as a New Epistemology for Logic? *Erkenntnis*.

Becker Arenhart, J. R. (2022b). Logical Anti-Exceptionalism meets the "Logic-as-Models" Approach. *Theoria*.

Berger, A. (2011). Kripke on the Incoherency of Adopting a Logic. In Berger, A., editor, *Saul Kripke*. Cambridge University Press.

Berker, S. (2013). The Rejection of Epistemic Consequentialism. *Philosophical Issues*, 23:363–387.

Berto, F. (2011). *There's Something About Gödel: The Complete Guide to the Incompleteness Theorem*. John Wiley & Sons.

Besong, B. (2014). Moral Intuitionism and Disagreement. *Synthese*, 191:2767–789.

Besson, C. (2019). Knowledge of Logical Generality and the Possibility of Deductive Reasoning. In Chan, T. & Nes, A., editor, *Inference and Consciousness*. Routledge.

Boghossian, P. and Peacocke, C. (2000). *New Essays on the A Priori*. Oxford University Press.

Bonevac, D. (2018). Defaulting on Reasons. *Noûs*, 52(2):229–259.

BonJour, L. (1985). *The Structure of Empirical Knowledge*. Harvard University Press.

BonJour, L. (1998). *In Defense of Pure Reason: A Rationalist Account of A Priori Justification*. Cambridge University Press.

Bradley, D. (2021). Uniqueness and Modesty: How Permissivists Can Live on the Edge. *Mind*, 130(520):1087–1098.

Brewka, G. (1989). Preferred Subtheories: An Extended Logical Framework for Default Reasoning. In Sridharan, N., editor, *Proceedings of the eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 1043–1048. Morgan Kaufmann Publishers.

Brewka, G. (1994a). Adding Priorities and Specificity to Default Logic. In *Logics in Artificial Intelligence: European Workshop JELIA 94*, pages 247–260. Springer.

Brewka, G. (1994b). Reasoning about Priorities in Default Logic. In *AAAI*, volume 1994, pages 940–945.

Brewka, G. and Eiter, T. (2000). Prioritizing Default Logic. In Hölldobler, S., editor, *Intellectics and Computational Logic*, volume 19 of *Applied Logic Series*, pages 27–45.

Broncano-Berrocal, F. and Simion, M. (2021). Disagreement and Epistemic Improvement. *Synthese*, 199(5-6):14641–14665.

Broome, J. (2007). Wide or Narrow Scope? *Mind*, 116(462):359–370.

Brouwer, L. (1975). Intuitionism and Formalism. In Heyting, A., editor, *Philosophy and Foundations of Mathematics*, pages 123–138. Elsevier North-Holland.

Brown, J. (2018). *Fallibilism: Evidence and Knowledge*. Oxford University Press.

Burge, T. (2003). Perceptual Entitlement. *Philosophy and Phenomenological Research*, 67(3):503–548.

Burge, T. and Peacocke, C. (1996). Our Entitlement to Self-Knowledge: II. Christopher Peacocke: Entitlement, Self-Knowledge and Conceptual Redeployment. In *Proceedings of the Aristotelian Society*, volume 96, pages 117–158.

Carlson, M. (2022). Anti-Exceptionalism and the Justification of Basic Logical Principles. *Synthese*, 200(3):214.

Carnap, R. (2014). *Logical Syntax of Language*. Routledge.

Carr, J. R. (2021). Why Ideal Epistemology? *Mind*, 131(524):1131–1162.

Carrol, L. (1895). What the Tortoise Said to Achilles. *Mind*, 4(14):278–280.

Carter, J. A. and Bondy, P. (2019). *Well-Founded Belief: New Essays on the Epistemic Basing Relation*. Routledge.

Chalmers, D. J. (2011). Verbal Disputes. *Philosophical Review*, 120(4):515–566.

Chislenko, E. (2021). How Can Belief Be Akratic? *Synthese*, 199:13925–13948.

Chow, T. Y. (2019). The Consistency of Arithmetic. *The Mathematical Intelligencer*, 41(1):22–30.

Christensen, D. (2007). Epistemology of Disagreement: The Good News. *The Philosophical Review*, 116(2):187–217.

Christensen, D. (2009). Disagreement as Evidence: The Epistemology of Controversy. *Philosophy Compass*, 4(5):756–767.

Christensen, D. (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research*, 81(1):185–215.

Christensen, D. (2011). Disagreement, Question-Begging and Epistemic Self-Criticism. *Philosophers Imprint*, 11(6):1–22.

Christensen, D. (2014). Disagreement and Public Controversy. In Lackey, J., editor, *Essays in Collective Epistemology*, pages 142–164. Oxford University Press.

Christensen, D. (2016). Conciliation, Uniqueness and Rational Toxicity. *Noûs*, 50(3):584–603.

Christensen, D. (2019). Formulating Independence. In Skipper, M. and Steglich-Petersen, A., editors, *Higher-Order Evidence. New Essays*, pages 13–34. Oxford University Press.

Christensen, D. (2021). Akratic (Epistemic) Modesty. *Philosophical Studies*, 178(7):2191–2214.

Christensen, D. (2022). Epistemic Akrasia: No Apology Required. *Noûs*.

Christensen, D. and Lackey, J. (2013). *The Epistemology of Disagreement: New Essays*. Oxford University Press.

Chudnoff, E. (2011). What Intuitions Are Like. *Philosophy and Phenomenological Research*, 82(3):625–654.

Cieśliński, C. (2017). *The Epistemic Lightness of Truth: Deflationism and its Logic*. Cambridge University Press.

Cohen, S. (2013). A Defense of the (Almost) Equal Weight View. In Christensen, D. and Lackey, J., editors, *The Epistemology of Disagreement: New Essays*, pages 98–117. Oxford University Press.

Cohnitz, D. (2020). Verbal Disputes and Deep Conceptual Disagreements. *Trames*, 24(3):279–294.

Cohnitz, D. and Estrada-González, L. (2019). *An Introduction to the Philosophy of Logic*. Cambridge University Press.

Coliva, A. (2010). *Moore and Wittgenstein: Scepticism, Certainty and Common Sense*. Springer.

Coliva, A. and Moyal-Sharrock, D. (2016). *Hinge Epistemology*. Brill.

Conee, E. (1987). Evident, but Rationally Unacceptable. *Australasian Journal of Philosophy*, 65(3):316–326.

Conee, E. (2010). Rational Disagreement Defended. In Feldman, R. and Warfield, T. A., editors, *Disagreement*, pages 69–90. Oxford University Press.

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

Constantin, J. and Grundmann, T. (2020). Epistemic Authority: Preemption Through Source Sensitive Defeat. *Synthese*, 197(9):4109–4130.

Cotnoir, A. J. (2018). Logical Nihilism. In Jeremy Wyatt, N. J. L. L. P. and Kellen, N., editors, *Pluralisms in Truth and Logic*, pages 301–329. Palgrave Macmillan.

Daniels, N. (1979). Wide Reflective Equilibrium and Theory Acceptance in Ethics. *The Journal of Philosophy*, 76(5):256–282.

Daniels, N. (1996). *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge University Press.

Daoust, M. (2019). Epistemic Akrasia and Epistemic Reasons. *Episteme*, 16(3):282–302.

Davidson, D. (2001). How Is Weakness of the Will Possible? In *Essays on Actions and Events*. Oxford University Press.

Decker, J. (2014). Conciliation and Self-Incrimination. *Erkenntnis*, 79(5):1099–1134.

Delgrande, J. P., Schaub, T., Tompits, H., , and Wang, K. (2004). A Classification and Survey of Preference Handling Approaches in Nonmonotonic Reasoning. *Computational Intelligence*, 20:308–334.

Devitt, M. and Roberts, J. R. (202X). Changing our Logic: A Quinean Perspective. *Mind*.

Dietrich, F. and Spiekermann, K. (2022). Jury Theorems. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2022 edition.

Douven, I. (2021). Abduction. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition.

Dretske, F. (1971). Conclusive Reasons. *Australasian Journal of Philosophy*, 49(1):1–22.

Dretske, F. (2000). Entitlement: Epistemic Rights without Epistemic Duties? *Philosophy and Phenomenological Research*, 60(3):591–606.

Dummett, M. (1991). *The Logical Basis of Metaphysics*. Harvard University Press.

Dutilh Novaes, C. (2012). *Formal Languages in Logic: A Philosophical and Cognitive Analysis*. Cambridge University Press.

Dutilh Novaes, C. (2015). A Dialogical, Multi-Agent Account of the Normativity of Logic. *Dialectica*, 69(4):587–609.

Dutilh Novaes, C. (2020). *The Dialogical Roots of Deduction: Historical, Cognitive, and Philosophical Perspectives on Reasoning*. Cambridge University Press.

Elga, A. (2005). On Overrating Oneself... and Knowing It. *Philosophical Studies*, 123(1):115–124.

Elga, A. (2007). Reflection and Disagreement. *Noûs*, 41:478–502.

Elga, A. (2010). How to Disagree about How to Disagree. In Feldman, R. and Warfield, T., editors, *Disagreement*, pages 175–186. Oxford University Press.

Engel, P. (2016). Epistemic Norms and the Limits of Epistemology. *International Journal for the Study of Skepticism*, 6(2-3):228–247.

Enoch, D. (2010). Not Just a Truthometer: Taking Oneself Seriously (but Not Too Seriously) in Cases of Peer Disagreement. *Mind*, 119(476):953–997.

Etherington, D. and Reiter, R. (1983). On Inheritance Hierarchies with Exceptions. In *Proceedings of the Third National Conference on Artificial Intelligence (AAAI-83)*, pages 104–108.

Evershed, J. (2021). Another Way Logic Might Be Normative. *Synthese*, 199(3-4):5861–5881.

Fassio, D. and Logins, A. (2023). Justification and Gradability. *Philosophical Studies*.

Feldman, R. (2005a). Deep Disagreement, Rational Resolutions, and Critical Thinking. *Informal Logic*, 25(1).

Feldman, R. (2005b). Respecting the Evidence. *Philosophical Perspectives*, 19:95–119.

Feldman, R. (2006). Epistemological Puzzles about Disagreement. In Hetherington, S., editor, *Epistemology Futures*, pages 216–236. Oxford University Press.

Feldman, R. (2009). Evidentialism, Higher-Order Evidence, and Disagreement. *Episteme*, 6(3):294–312.

Feldman, R. and Conee, E. (1985). Evidentialism. *Philosophical Studies*, 48(1):15–34.

Feldman, R. and Warfield, T. A. (2010). *Disagreement*. Oxford University Press.

Ferrari, F., Martin, B., and Fogliani Sforza, M. P. (2023). Anti-Exceptionalism about Logic: An Overview. *Synthese*, 201(2):62.

Field, C. (2019). It's Ok to Make Mistakes: Against the Fixed Point Thesis. *Episteme*, 16(2):175–185.

Field, C. (2021). Giving Up the Enkratic Principle. *Logos & Episteme*, 12(1):7–28.

Field, H. (2015). What is Logical Validity. In Caret, C. R. and Hjortland, O. T., editors, *Foundations of Logical Consequence*, pages 33–70. Oxford University Press.

Field, H. (2017). Disarming a Paradox of Validity. *Notre Dame Journal of Formal Logic*, 58(1):1–19.

Finn, S. (2019). The Adoption Problem and Anti-Exceptionalism about Logic. *Australasian Journal of Logic*, 16(7):231–249.

Fischer, M., Horsten, L., and Nicolai, C. (2021). Hypatia's Silence: Truth, Justification, and Entitlement. *Noûs*, 55(1):62–85.

Fogal, D. and Worsnip, A. (2021). Which Reasons? Which Rationality? *Ergo*, 8(11).

Fogelin, R. (2005). The Logic of Deep Disagreements. *Informal Logic*, 25(1).

Foley, R. (2001). *Intellectual Trust in Oneself and Others*. Cambridge University Press.

Frances, B. (2008). Live Skeptical Hypotheses. In Greco, J., editor, *Oxford Handbook of Skepticism*, pages 225–245. Oxford University Press.

Frances, B. (2014). *Disagreement*. Polity Press.

Frances, B. and Matheson, J. (2019). Disagreement. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2019 edition.

Fraser, R. (2022). Mushy Akrasia: Why Mushy Credences Are Rationally Permissible. *Philosophy and Phenomenological Research*, 105:79–106.

Frege, G. (1948). Sense and Reference. *The Philosophical Review*, 57(3):209–230.

Frege, G. (1956). The Thought: A Logical Inquiry. *Mind*, 65(259):289–311.

Frege, G. (1997). Logic. In Beaney, M., editor, *The Frege Reader*. Blackwell.

Frege, G. (2013). *Gottlob Frege: Basic Laws of Arithmetic*, volume 1. Oxford University Press.

Friemann, R. (2005). Emotional Backing and the Feeling of Deep Disagreement. *Informal Logic*, 25(1).

Gentzen, G. (1936). Die Widerspruchsfreiheit der Reinen Zahlentheorie. *Mathematische Annalen*, 112(1):493–565.

Ginsberg, M., editor (1987). *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann Publishers.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38(1):173–198.

Gödel, K. (1964). What is Cantor's Continuum Problem? In Benacerraf, P. J. S., editor, *Philosophy of Mathematics: Selected Readings*, pages 259–273. Cambridge University Press.

Goldman, A. and Whitcomb, D. (2011). *Social Epistemology: Essential Readings*. Oxford University Press.

Goldman, A. I. (1979). What is Justified Belief? In Pappas, G. S., editor, *Justification and Knowledge: New Studies in Epistemology*, pages 1–23. Springer.

Goldman, A. I. (1986). *Epistemology and Cognition*. Harvard University Press.

Goldman, A. I. (2001). Experts: Which Ones Should You Trust? *Philosophy and Phenomenological Research*, 63(1):85–110.

Goldman, A. I. (2010). Epistemic Relativism and Reasonable Disagreement. In Feldman, R. and Warfield, T. A., editors, *Disagreement*, pages 187–215.

Goodman, N. (1983). *Fact, Fiction, and Forecast*. Harvard University Press.

Greco, D. (2014). A Puzzle about Epistemic Akrasia. *Philosophical Studies*, 167(2):201–219.

Griffiths, O. and Paseau, A. (2022). *One True Logic*. Oxford University Press.

Grundmann, T. (2019). How to Respond Rationally to Peer Disagreement: The Preemption View. *Philosophical Issues*, 29(1):129–142.

Gutting, G. (1982). *Religious Belief and Religious Skepticism*. University of Notre Dame Press.

Haack, S. (1974). *Deviant Logic: Some Philosophical Issues*. Cambridge University Press.

Hales, S. D. (2014). Motivations for Relativism as a Solution to Disagreements. *Philosophy*, 89(1):63–82.

Hansson, S. O. (2017). Logic of Belief Revision. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition.

Harman, G. (1984). Logic and Reasoning. *Synthese*, 60(1):107–127.

Harman, G. (1986). *Change in View: Principles of Reasoning*. MIT Press.

Harris, J. H. (1982). What's so Logical about the "Logical" Axioms? *Studia Logica*, 41:159–171.

Hattiangadi, A. (2018). Logical Disagreement. In McHugh, C., editor, *Metaepistemology*, pages 88–106. Oxford University Press.

Hattiangadi, A. and Andersen, F. J. (202X). Logical Disagreement. In *The Oxford Handbook of Philosophy of Logic*. Oxford University Press.

Hawthorne, J. and Logins, A. (2021). Graded Epistemic Justification. *Philosophical Studies*, 178:1845–1858.

Hazlett, A. (2014). Entitlement and Mutually Recognized Reasonable Disagreement. *Episteme*, 11(1):1–25.

Henderson, L. (2022). Higher-Order Evidence and Losing One's Conviction. *Noûs*, 56(3):513–529.

Hintikka, K. J. J. (2005). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. College Publications.

Hjortland, O. T. (2014). Verbal Disputes in Logic: Against Minimalism for Logical Connectives. *Logique et Analyse*.

Hjortland, O. T. (2017). Anti-Exceptionalism about Logic. *Philosophical Studies*, 174(3):631–658.

Hjortland, O. T. (2019). What Counts as Evidence for a Logical Theory? *The Australasian Journal of Logic*, 16(7):250–282.

Hjortland, O. T. (2022). Disagreement about Logic. *Inquiry*, 65(6):660–682.

Hlobil, U. (2021). Limits of Abductivism about Logic. *Philosophy and Phenomenological Research*, 103(2):320–340.

Horowitz, S. (2014). Epistemic Akrasia. *Noûs*, 48(4):718–744.

Horowitz, S. (2022). Higher-Order Evidence. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2022 edition.

Horsten, L. and Leigh, G. E. (2017). Truth is Simple. *Mind*, 126(501):195–232.

Horty, J. F. (2007a). Defaults with Priorities. *Journal of Philosophical Logic*, 36(4):367–413.

Horty, J. F. (2007b). Reasons as Defaults. *Philosophers' Imprint*, 7:1–28.

Horty, J. F. (2012). *Reasons as Defaults*. Oxford University Press.

Horty, J. F. (2016). Reasoning with Precedents as Constrained Natural Reasoning. In Lord, E. and Maguire, B., editors, *Weighing Reasons*. Oxford University Press.

Hume, D. (1975). *An Enquiry into the Human Understanding*. Oxford University Press.

Jackson, E. and Tan, P. (2022). Epistemic Akrasia and Belief-Credence Dualism. *Philosophy and Phenomenological Research*, 104(3):717–727.

Jeffrey, R. C. and Burgess, J. P. (2006). *Formal Logic: Its Scope and Limits*. Hackett Publishing.

Jenkins, C. S. (2007). Entitlement and Rationality. *Synthese*, 157:25–45.

Jenkins, C. S. (2014). Merely Verbal Disputes. *Erkenntnis*, 79(1):11–30.

Jønch-Clausen, K. and Kappel, K. (2015). Social Epistemic Liberalism and the Problem of Deep Epistemic Disagreements. *Ethical Theory and Moral Practice*, 18(2):371–384.

Kahan, D. M. and Braman, D. (2006). Cultural Cognition and Public Policy. *Yale Law & Policy Review*, 24:149–172.

Kahneman, D., Slovic, S. P., Slovic, P., and Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kant, I. (2003). *Critique of Pure Reason*. Dover.

Kappel, K. (2012). The Problem of Deep Disagreement. *Discipline Filosofiche*, 22(2):7–25.

Kappel, K. (2019a). Bottom Up Justification, Asymmetric Epistemic Push, and the Fragility of Higher Order Justification. *Episteme*, 16(2):119–138.

Kappel, K. (2019b). Escaping the Akratic Trilemma. In Skipper, M. and Steglich-Petersen, A., editors, *Higher-Order Evidence. New Essays*, pages 124–143. Oxford University Press.

Kappel, K. (2021). Higher Order Evidence and Deep Disagreement. *Topoi*, 40(5):1039–1050.

Kappel, K. and Andersen, F. J. (2019). Moral Disagreement and Higher-Order Evidence. *Ethical Theory and Moral Practice*, 22(5):1103–1120.

Kauss, D. (2023). A Rational Agent With Our Evidence. *Erkenntnis*.

Kearl, T. (2020). Epistemic Akrasia and Higher-Order Beliefs. *Philosophical Studies*, 177(9):2501–2515.

Kelly, T. (2005). The Epistemic Significance of Disagreement. In Hawthorne, J. and Gendler, T., editors, *Oxford Studies in Epistemology, I*, pages 167–196. Oxford University Press.

Kelly, T. (2008). Disagreement, Dogmatism, and Belief Polarization. *The Journal of Philosophy*, 105(10):611–633.

Kelly, T. (2010). Peer Disagreement and Higher-Order Evidence. In Feldman, R. and Warfield, T. A., editors, *Disagreement*, pages 111–174. Oxford University Press.

Kelly, T. (2013). Disagreement and the Burdens of Judgment. In Christensen, D. and Lackey, J., editors, *The Epistemology of Disagreement: New Essays*, pages 31–53. Oxford University Press.

Kelly, T. (2014). Evidence Can Be Permissive. In Steup, M., Turri, J., and Sosa, E., editors, *Contemporary Debates in Epistemology*, pages 298–311. Wiley Blackwell.

Kelp, C. (2023). *The Nature and Normativity of Defeat*. Cambridge University Press.

King, N. L. (2012). Disagreement: What's the Problem? Or a Good Peer is Hard to Find. *Philosophy and Phenomenological Research*, 85(2):249–272.

Knoks, A. (2021a). Conciliatory Reasoning, Self-Defeat, and Abstract Argumentation. *The Review of Symbolic Logic*.

Knoks, A. (2021b). Misleading Higher-Order Evidence, Conflicting Ideals, and Defeasible Logic. *Ergo*, 8(6):141–174.

Knoks, A. (2022). Conciliatory Views, Higher-Order Disagreements, and Defeasible Logic. *Synthese*, 200(2):1–23.

Koksvik, O. (2017). The Phenomenology of Intuition. *Philosophy Compass*, 12(1):e12387.

Kolodny, N. and Brunero, J. (2023). Instrumental Rationality. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2023 edition.

Koons, R. (2017). Defeasible Reasoning. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition.

Kopec, M. and Titelbaum, M. G. (2016). The Uniqueness Thesis. *Philosophy Compass*, 11(4):189–200.

Korcz, K. A. (2021). The Epistemic Basing Relation. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2021 edition.

Kripke, S. (1974). Princeton Lectures on the Nature of Logic. *Unpublished*.

Kripke, S. (1976). Outline of a Theory of Truth. *The Journal of Philosophy*, 72(19):690–716.

Lackey, J. (2010). A Justificationalist View of Disagreement's Epistemic Significance. In Adrian Haddock, A. M. and Pritchard, D., editors, *Social Epistemology*, pages 298–325. Oxford University Press.

Lackey, J. (2013). Disagreement and Belief Dependence: Why Numbers Matter. In Christensen, D. and Lackey, J., editors, *The Epistemology of Disagreement: New Essays*, pages 243–268. Oxford University Press.

Lagewaard, T. (2021). Epistemic Injustice and Deepened Disagreement. *Philosophical Studies*, 178(5):1571–1592.

Lasonen-Aarnio, M. (2014). Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research*, 88(2):314–345.

Lasonen-Aarnio, M. (2020). Enkrasia or Evidentialism? Learning to Love Mismatch. *Philosophical Studies*, 177(3):597–632.

Lewis, D. (1979). Scorekeeping in a Language Game. *Journal of Philosophical Journal*, 8:539–559.

Lewis, D. (2004). Letters to Beall and Priest. In Graham Priest, J. B. and Armour-Garb, B., editors, *The Law of Non-Contradiction: New Philosophical Essays*, pages 176–177. Oxford University Press.

Littlejohn, C. (2012). *Justification and the Truth-Connection*. Cambridge University Press.

Littlejohn, C. (2018). Stop Making Sense? On a Puzzle about Rationality. *Philosophy and Phenomenological Research*, 96(2):257–272.

Littlejohn, C. (2020). Should we be Dogmatically Conciliatory? *Philosophical Studies*, 177(5):1381–1398.

Lord, E. (2014). From Independence to Conciliationism: An Obituary. *Australasian Journal of Philosophy*, 92(2):365–377.

Lynch, M. (2010). Epistemic Circularity and Epistemic Incommensurability. In Haddock A, Millar A, P. D., editor, *Social Epistemology*, pages 262–277. Oxford University Press.

Lynch, M. P. (2016). After the Spade Turns: Disagreement, First Principles and Epistemic Contractarianism. *International Journal for the Study of Skepticism*, 6(2-3):248–259.

MacFarlane, J. (2000). What Does It Mean To Say That Logic Is Formal. *PhD Thesis, University of Pittsburgh.*

MacFarlane, J. (2004). In What Sense (if any) is Logic Normative for Thought? *Unpublished*.

Maddy, P. (2014). *The Logical Must: Wittgenstein on Logic*. Oxford University Press.

Makinson, D. (1994). General Patterns in Non-Monotonic Reasoning. In Gabbay, D. M., Hogger, C. J., and Robinson, J. A., editors, *Handbook of Logic in Artificial Intelligence and Logic Programming: Nonmonotonic Reasoning and Uncertain Reasoning (vol. 3)*, pages 35–110. Oxford University Press.

Makinson, D. and Gärdenfors, P. (1991). Relations between the Logic of Theory Change and Nonmonotonic Logic. In *The Logic of Theory Change*, pages 183–205. Springer.

Makinson, D. C. (1965). The Paradox of the Preface. *Analysis*, 25(6):205–207.

Martin, B. (2021a). Anti-Exceptionalism about Logic and the Burden of Explanation. *Canadian Journal of Philosophy*, 51(8):602–618.

Martin, B. (2021b). Identifying Logical Evidence. *Synthese*, 198(10):9069–9095.

Martin, B. (2021c). Searching for Deep Disagreement in Logic: The Case of Dialetheism. *Topoi*, 40(5):1127–1138.

Martin, B. (2022). The Philosophy of Logical Practice. *Metaphilosophy*, 53(2-3):267–283.

Martin, B. and Hjortland, O. (2021). Logical Predictivism. *Journal of Philosophical Logic*, 50:285–318.

Martin, B. and Hjortland, O. T. (2022). Anti-Exceptionalism about Logic as Tradition Rejection. *Synthese*, 200(2):148.

Martin, B. and Hjortland, O. T. (202X). Evidence in Logic. In Lasonen-Aarnio, M. and Littlejohn, C., editors, *Routledge Handbook of the Philosophy of Evidence*. Routledge.

Matheson, J. (2009). Conciliatory Views of Disagreement and Higher-Order Evidence. *Episteme*, 6(3):269–279.

Matheson, J. (2011). The Case for Rational Uniqueness. *Logos & Episteme*, 2(3):359–373.

Matheson, J. (2014). A Puzzle about Disagreement and Rationality. *Social Epistemology Review and Reply Collective*, 3(4).

Matheson, J. (2015). *The Epistemic Significance of Disagreement*. Springer.

Maudlin, T. (2005). The Tale of Quantum Logic. In Ben-Menahem, Y., editor, *Hilary Putnam*, pages 156–187. Cambridge University Press.

McCain, K. (2012). The Interventionist Account of Causation and the Basing Relation. *Philosophical Studies*, 159:357–382.

McCain, K. (2014). *Evidentialism and Epistemic Justification*. Routledge.

McCarthy, J. (1980). Circumscription – A Form of Non-Monotonic Reasoning. *Artificial Intelligence*, 13(1-2):27–39.

McCarthy, J. (1986). Applications of Circumscription to Formalizing Common Sense Knowledge. *Artificial Intelligence*, 28(1):89–116.

McDermot, D. and Doyle, J. (1980). Non-Monotonic Logic 1. *Artificial Intelligence*, 13:41–72.

McGinn, M. (1989). *Sense and Certainty: A Dissolution of Scepticism*. Wiley-Blackwell.

McKenna, R. (2023). *Non-Ideal Epistemology*. Oxford University Press.

Melchior, G. (202X). Rationally Irresolvable Disagreement. *Philosophical Studies*.

Meyer, R. K. (1985). Proving Semantical Completeness 'Relevantly' for R. In *Technical Report*. Australian National University.

Milne, P. (2009). II–Peter Milne: What is the Normative Role of Logic? In *Aristotelian Society Supplementary Volume*, volume 83, pages 269–298. Oxford University Press.

Mitova, V. (2018). *The Factive Turn in Epistemology*. Cambridge University Press.

Moon, A. (2018). Independence and New Ways to Remain Steadfast in the Face of Disagreement. *Episteme*, 15(1):65–79.

Moore, R. C. (1985). Semantical Considerations on Nonmonotonic Logic. *Artificial Intelligence*, 25(1):75–94.

Mortensen, C. (2013). *Inconsistent Mathematics*. Kluwer Academic Publishers.

Moyal-Sharrock, D. (2004). Introduction: The Idea of a Third Wittgenstein. In Moyal-Sharrock, D., editor, *The Third Wittgenstein*. Routledge.

Mulligan, T. (2015). Disagreement, Peerhood, and Three Paradoxes of Conciliationism. *Synthese*, 192(1):67–78.

Neta, R. (2019). The Basing Relation. *Philosophical Review*, 128(2):179–217.

Nolt, J. (2021). Free Logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2021 edition.

Nozick, R. (1983). *Philosophical Explanations*. Harvard University Press.

Oberauer, K. and Lewandowsky, S. (2019). Addressing the Theory Crisis in Psychology. *Psychonomic Bulletin & Review*, 26:1596–1618.

Owens, D. (2002). Epistemic Akrasia. *The Monist*, 85(3):381–397.

Padro, R. (2015). *What the Tortoise said to Kripke: The Adoption Problem and the Epistemology of Logic*. CUNY Academic Works.

Pappas, G. (2023). Internalist vs. Externalist Conceptions of Epistemic Justification. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2023 edition.

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

Parfit, D. (2011). *On What Matters*, volume 1-3. Oxford University Press.

Pettit, P. (2006). When to Defer to Majority Testimony – and When Not. *Analysis*, 66:179–87.

Plantinga, A. (2000). Pluralism: A Defense of Religious Exclusivism. In Quinn, P. L. and Meeker, K., editors, *The Philosophical Challenge of Religious Diversity*, pages 172–192. Oxford University Press.

Pollock, J. (1970). The Structure of Epistemic Justification. *American Philosophical Quarterly (Monograph Series 4)*, pages 62–78.

Pollock, J. (1974). *Knowledge and Justification*. Princeton University Press.

Pollock, J. (1984). Reliability and Justified Belief. *Canadian Journal of Philosophy 14*, pages 103–114.

Pollock, J. (1986). *Contemporary Theories of Knowledge*. Rowman and Littlefield.

Pollock, J. L. (1994). Justification and Defeat. *Artificial Intelligence*, 67(2):377–407.

Poole, D. (1988). A Logical Framework for Default Reasoning. *Artificial Intelligence*, 36:27–47.

Prawitz, D. (1965). *Natural Deduction: A Proof-Theoretical Study*. Dover.

Prawitz, D. (2006). Meaning Approached via Proofs. *Synthese*, 148:507–524.

Priest, G. (2005). *Doubt Truth to be a Liar*. Oxford University Press.

Priest, G. (2006). *In Contradiction*. Oxford University Press.

Priest, G. (2008). *An Introduction to Non-Classical Logic*. Cambridge University Press.

Priest, G. (2014). Revising Logic. In Rush, P., editor, *The Metaphysics of Logic*, pages 211–223. Cambridge University Press.

Priest, G. (2021). Logical Abductivism and Non-Deductive Inference. *Synthese*, 199:3207–3217.

Prior, A. N. (1960). The Runabout Inference-Ticket. *Analysis*, 21(2):38–39.

Pritchard, D. (2005). *Epistemic Luck*. Oxford University Press.

Pritchard, D. (2010). Epistemic Relativism, Epistemic Incommensurability and Wittgensteinian Epistemology. In Hales, S., editor, *Blackwell Companion to Relativism*, pages 266–285.

Pritchard, D. (2021). Wittgensteinian Hinge Epistemology and Deep Disagreement. *Topoi*, 40(5):1117–1125.

Putnam, H. (1969). Is Logic Empirical? In *Boston Studies in the Philosophy of Science: Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968*, pages 216–241. Springer.

Quine, W. v. O. (1951). Two Dogmas of Empiricism. *Philosophical Review*, 60(1):20–43.

Quine, W. v. O. (1953). Two Dogmas of Empiricism. In *From a Logical Point of View*. Harvard University Press.

Quine, W. v. O. (1986). *Philosophy of Logic*. Harvard University Press.

Quine, W. v. O. (2004). Truth by Convention. In *Quintessence*. Harvard University Press.

Quine, W. v. O. (1960). Carnap and Logical Truth. *Synthese*, 12:350–374.

Ranalli, C. (2020). Deep Disagreement and Hinge Epistemology. *Synthese*, 197(11):4975–5007.

Ranalli, C. (2021). What is Deep Disagreement? *Topoi*, 40:983–998.

Ranalli, C. and Lagewaard, T. (2022a). Deep Disagreement (Part 1): Theories of Deep Disagreement. *Philosophy Compass*, 17(12):e12886.

Ranalli, C. and Lagewaard, T. (2022b). Deep Disagreement (Part 2): Epistemology of Deep Disagreement. *Philosophy Compass*, 17(12):e12887.

Rawls, J. (1951). Outline of a Decision Procedure for Ethics. *The Philosophical Review*, 60(2):177–197.

Rawls, J. (2020). *A Theory of Justice*. Harvard University Press.

Read, S. (2006). Monism: The One True Logic. In DeVidi, D. and Kenyon, T., editors, *A Logical Approach to Philosophy: Essays in Honour of Graham Solomon*, pages 193–209. Springer.

Reiss, J. and Sprenger, J. (2020). Scientific Objectivity. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2020 edition.

Reiter, R. (1980). A Logic for Default Reasoning. *Artificial Intelligence*, 13:81–132.

Reiter, R. and Criscuolo, G. (1981). On Interacting Defaults. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI-81)*, pages 270–276.

Rendsvig, R. and Symons, J. (2019). Epistemic Logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition.

Resnik, M. (1985). Logic: Normative or Descriptive? The Ethics of Belief or a Branch of Psychology? *Philosophy of Science*, 52(2):221–238.

Resnik, M. (1996). Ought There to be but One Logic. In Copeland, B. J., editor, *Logic and Reality: Essays on the Legacy of Arthur Prior*, pages 489–517. Oxford University Press.

Resnik, M. (2004). Revising Logic. In Priest, G., Beall, J. C., and Armour-Garb, B., editors, *The Law of Non-Contradiction*, pages 178–193. Oxford University Press.

Restall, G. (2022). *Proofs and Models in Philosophical Logic*. Cambridge University Press.

Restall, G. (202X). *Proof, Rules and Meaning*.

Restall, G. and Standefer, S. (2023). *Logical Methods*. MIT Press.

Ribeiro, B. (2011). Epistemic Akrasia. *International Journal for the Study of Skepticism*, 1:18–25.

Ricketts, T. G. (1985). Frege, the Tractatus, and the Logocentric Predicament. *Noûs*, 19(1):3–15.

Rintanen, J. (1995). On Specificity in Default Logic. In Mellish, C., editor, *IJCAI'95*, pages 1474–1479. AAAI Press.

Rintanen, J. (1998). Lexicographic Priorities in Default Logic. *Artificial Intelligence*, 106(2):221–265.

Rorty, A. (1983). Akratic Believers. *American Philosophical Quarterly*, 20(2):175–183.

Rosa, L. (2012). Justification and the Uniqueness Thesis. *Logos & Episteme*, 3(4):571–577.

Rosa, L. (2016). Justification and the Uniqueness Thesis Again: A Response to Anantharaman. *Logos & Episteme*, 7(1):95–100.

Rosen, G. (2001). Nominalism, Naturalism, Epistemic Relativism. *Philosophical Perspectives*, 15:69–91.

Ross, R. (2021). Alleged Counterexamples to Uniqueness. *Logos & Episteme*, 12(2):203–213.

Rossberg, M. and Shapiro, S. (2021). Logic and Science: Science and Logic. *Synthese*, 199(3-4):6429–6454.

Rossi, A. (2023). From Collapse Theorems to Proof-Theoretic Arguments. *Australasian Journal of Logic*, 20(1).

Roush, S. (2017). Epistemic Self-Doubt. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2017 edition.

Russell, B. (1919). *Introduction to Mathematical Philosophy*. George Allen & Unwin.

Russell, B. (2009). *The Philosophy of Logical Atomism*. Routledge.

Russell, G. (2014). Metaphysical Analyticity and the Epistemology of Logic. *Philosophical Studies*, 171:161–175.

Russell, G. (2015). The Justification of the Basic Laws of Logic. *Journal of Philosophical Logic*, 44:793–803.

Russell, G. (2017). An Introduction to Logical Nihilism. In *Logic, Methodology and Philosophy of Science–Proceedings of the 15th International Congress*, pages 125–135. College Publications.

Russell, G. (2018a). Logical Nihilism: Could There Be No Logic? *Philosophical Issues*, 28(1):308–324.

Russell, G. (2018b). Varieties of Logical Consequence by Their Resistance to Logical Nihilism. In Jeremy Wyatt, N. J. L. L. P. and Kellen, N., editors, *Pluralisms in Truth and Logic*, pages 331–361. Palgrave Macmillan.

Russell, G. (2019a). Deviance and Vice: Strength as a Theoretical Virtue in the Epistemology of Logic. *Philosophy and Phenomenological Research*, 99(3):548–563.

Russell, G. (2019b). Logical Pluralism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition.

Russell, G. (2020). Logic isn't Normative. *Inquiry*, 63(3-4):371–388.

Sagi, G. (2021). Logic as a Methodological Discipline. *Synthese*, 199(3-4):9725–9749.

Schupbach, J. N. (2022). *Bayesianism and Scientific Reasoning*. Cambridge University Press.

Schurz, G. (2021). Meaning-Preserving Translations of Non-Classical Logics into Classical Logic: Between Pluralism and Monism. *Journal of Philosophical Logic*, 51:27–55.

Sellars, W. (1953). Inference and Meaning. *Mind*, 62(247):313–338.

Shapiro, S. (2000). The Status of Logic. In Boghossian, P. and Peacocke, C., editors, *New Essays on the A Priori*, pages 333–366. Oxford University Press.

Shapiro, S. (2014). *Varieties of Logic*. Oxford University Press.

Sider, T. (2010). *Logic for Philosophy*. Oxford University Press.

Sider, T. (2013). *Writing the Book of the World*. Oxford University Press.

Silva, P. (2023). Merely Statistical Evidence: When and Why it Justifies Belief. *Philosophical Studies*.

Skipper, M. (2019). Reconciling Enkrasia and Higher-Order Defeat. *Erkenntnis*, 84(6):1369–1386.

Skipper, M. (2021). Does Rationality Demand Higher-Order Certainty? *Synthese*, 198(12):11561–11585.

Skipper, M. and Steglich-Petersen, A. (2019). *Higher-Order Evidence. New Essays*. Oxford University Press.

Sliwa, P. and Horowitz, S. (2015). Respecting All the Evidence. *Philosophical Studies*, 172(11):2835–2858.

Smart, J. A. (2021). Disbelief is a Distinct Doxastic Attitude. *Synthese*, 198(12):11797–11813.

Smithies, D. (2012). Moore's Paradox and the Accessibility of Justification. *Philosophy and Phenomenological Research*, 85(2):273–300.

Sorensen, R. (1988). *Blindspots*. Oxford University Press.

Sorensen, R. (2020). Epistemic Paradoxes. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2020 edition.

Sosa, E. (1991). *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge University Press.

Sosa, E. (2010). The Epistemology of Disagreement. In Adrian Haddock, Alan Millar, D. P., editor, *Social Epistemology*, pages 278–297. Oxford University Press.

Srinivasan, A. and Hawthorne, J. (2013). Disagreement without Transparency: Some Bleak Thoughts. In Christensen, D. and Lackey, J., editors, *The Epistemology of Disagreement. New Essays*, pages 9–30. Oxford University Press.

Steinberger, F. (2019a). Consequence and Normative Guidance. *Philosophy and Phenomenological Research*, 98(2):306–328.

Steinberger, F. (2019b). Logical Pluralism and Logical Normativity. *Philosophers' Imprint*, 19(12).

Steinberger, F. (2019c). Three Ways in which Logic Might Be Normative. *The Journal of Philosophy*, 116(1):5–31.

Steinberger, F. (2022). The Normative Status of Logic. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2022 edition.

Strasser, C. and Antonelli, A. (2019). Non-Monotonic Logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition.

Sudduth, M. (2008). Defeaters in Epistemology. *Internet Encyclopedia of Philosophy*.

Tajer, D. (2022a). A Simple Solution to the Collapse Argument for Logical Pluralism. *Inquiry*, 0(0):1–18.

Tajer, D. (2022b). Anti-Exceptionalism and Methodological Pluralism in Logic. *Synthese*, 200(3):195.

Talbot, B. (2014). Truth Promoting Non-Evidential Reasons for Belief. *Philosophical Studies*, 168:599–618.

Talbott, W. (2016). Bayesian Epistemology. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2016 edition.

Tennant, N. (2007). Existence and Identity in Free Logic: A Problem for Inferentialism? *Mind*, 116(464):1055–1078.

Thomason, R. (2018). Logic and Artificial Intelligence. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition.

Titelbaum, M. G. (2015). Rationality Fixed Point (or: In Defense of Right Reason). In Gendler, T. S. and Hawthorne, J., editors, *Oxford Studies in Epistemology, Volume 5*, pages 253–294. Oxford University Press.

Titelbaum, M. G. (2019). Return to Reason. In Skipper, M. and Steglich-Petersen, A., editors, *Higher-Order Evidence. New Essays*, pages 226–245. Oxford University Press.

Titelbaum, M. G. and Kopec, M. (2019). When Rational Reasoners Reason Differently. In Jackson, M. B. and Jackson, B. B., editors, *Reasoning: New Essays on Theoretical and Practical Thinking*, pages 205–231. Oxford University Press.

Touretzky, D. S. (1986). *The Mathematics of Inheritance Systems*. Morgan Kaufmann Publishers.

Tversky, A. and Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481):453–458.

Väyrynen, P. (2021). Thick Ethical Concepts. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2021 edition.

Wason, P. C. (1968). Reasoning about a Rule. *Quarterly Journal of Experimental Psychology*, 20(3):273–281.

Weatherson, B. (2007). Disagreeing about Disagreement. *Unpublished Manuscript*.

Weatherson, B. (2010). Do Judgments Screen Evidence? *Unpublished Manuscript*.

Weber, Z., Badia, G., and Girard, P. (2016). What is an Inconsistent Truth Table? *Australasian Journal of Philosophy*, 94(3):533–548.

Wedgwood, R. (2007). *The Nature of Normativity*. Oxford University Press.

Wedgwood, R. (2010). The Moral Evil Demons. In Feldman, R. and Warfield, T. A., editors, *Disagreement*, pages 216–246. University Press Oxford.

Wedgwood, R. (2012). Justified Inference. *Synthese*, 189(2):273–295.

Wedgwood, R. (2019). Moral Disagreement and Inexcusable Irrationality. *American Philosophical Quarterly*, 56(1):97–108.

White, R. (2014). Evidence Cannot be Permissive. In Matthias Steup, John Turri, E. S., editor, *Contemporary Debates in Epistemology*, pages 312–323. Wiley Blackwell.

White, R. (2023). Evidence and Truth. *Philosophical Studies*, 180(3):1049–1057.

Williamson, T. (1988). Equivocation and Existence. In *Proceedings of the Aristotelian Society*, volume 88, pages 109–127.

Williamson, T. (1999). A Note on Truth, Satisfaction and the Empty Domain. *Analysis*, 59(1):3–8.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.

Williamson, T. (2007). *The Philosophy of Philosophy*. Wiley Blackwell.

Williamson, T. (2011a). Improbable Knowing. In Dougherty, T., editor, *Evidentialism and its Discontents*, pages 147–164. Oxford University Press.

Williamson, T. (2011b). Knowledge First Epistemology. In Sven Bernecker, D. P., editor, *The Routledge Companion to Epistemology*, pages 208–218. Routledge.

Williamson, T. (2013a). Gettier Cases in Epistemic Logic. *Inquiry*, 56(1):1–14.

Williamson, T. (2013b). Knowledge First. In Matthias Steup, John Turri, E. S., editor, *Contemporary Debates in Epistemology*, pages 1–9. Wiley Blackwell.

Williamson, T. (2013c). *Modal Logic as Metaphysics*. Oxford University Press.

Williamson, T. (2014). Very Improbable Knowing. *Erkenntnis*, 79(5):971–999.

Williamson, T. (2017a). Model-Building in Philosophy. In Blackford, R. and Broderick, D., editors, *Philosophy's Future*, pages 159–71. Wiley Blackwell.

Williamson, T. (2017b). Semantic Paradoxes and Abductive Methodology. In Armour-Garb, B., editor, *Reflections on the Liar*, pages 325–346. Oxford University Press.

Williamson, T. (2019). Evidence of Evidence in Epistemic Logic. In Skipper, M. and Steglich-Petersen, A., editors, *Higher-Order Evidence. New Essays*, pages 265–297. Oxford University Press.

Williamson, T. (2020a). *Philosophical Method: A Very Short Introduction*. Oxford University Press.

Williamson, T. (2020b). Vagueness: A Global Approach, by Kit Fine. *Mind*, 131(522):675–683.

Williamson, T. (202X). Accepting a Logic, Accepting a Theory. In Padro, R. and Weiss, Y., editors, *Saul Kripke on Modal Logic*. Springer.

Wittgenstein, L. (1969a). *The Blue and Brown Books*. Wiley Blackwell.

Wittgenstein, L. (1969b). *On Certainty*. Wiley Blackwell.

Wittgenstein, L. (2009). *Philosophical Investigations*. Wiley Blackwell.

Woods, J. (2019). Against Reflective Equilibrium for Logical Theorizing. *The Australasian Journal of Logic*, 16(7):319–341.

Woods, J. (2023). A Sketchy Logical Conventionalism. In *Aristotelian Society Supplementary Volume*, volume 97, pages 29–46. Oxford University Press.

Woodward, J. (2006). Some Varieties of Robustness. *Journal of Economic Methodology*, 13(2):219–240.

Worsnip, A. (2018). The Conflict of Evidence and Coherence. *Philosophy and Phenomenological Research*, 96(1):3–44.

Wright, C. (1986). Inventing Logical Necessity. In Butterfield, J., editor, *Language, Mind and Logic*, pages 187–209. Cambridge University Press.

Wright, C. (2004a). Intuition, Entitlement and the Epistemology of Logical Laws. *Dialectica*, 58(1):155–175.

Wright, C. (2004b). Warrant for Nothing (and Foundations for Free)? *Aristotelian Society Supplementary Volume*, 78(1):167–212.

Wright, C. (2021). Alethic Pluralism, Deflationism, and Faultless Disagreement. *Metaphilosophy*, 52(3-4):432–448.

Ye, R. (2020). Higher-Order Defeat and Intellectual Responsibility. *Synthese*, 197(12):5435–5455.

Ye, R. (2023). *Higher-Order Evidence and Calibrationism*. Cambridge University Press.

Zach, R. (2019). *Incompleteness and Computability: An Open Introduction to Gödel's Theorems*. Open Logic Project.

Zagzebski, L. T. (2012). *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. Oxford University Press.

Zanetti, L. (2021). Abstraction without Exceptions. *Philosophical Studies*, 178(10):3197–3216.